(2)

STUDIES OF GENETIC VARIATION IN THE AIDS VIRUS: RELEVANCE TO DISEASE PATHOGENESIS, ANTI-VIRAL THERAPY, AND VACCINE DEVELOPMENT

FINAL REPORT

GEORGE M. SHAW

JUNE 30, 1990

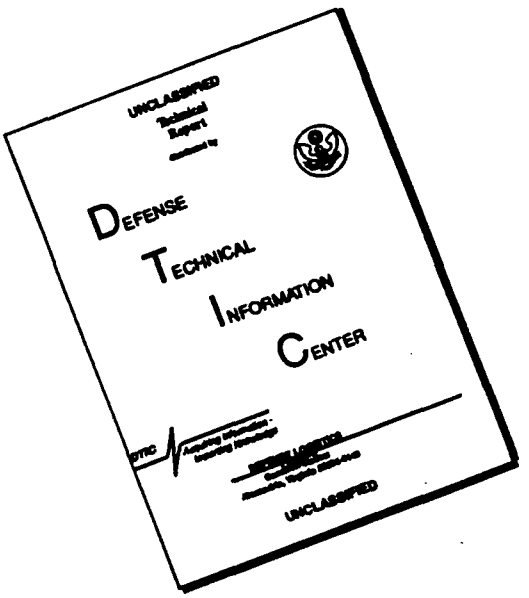
Supported by

U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Frederick, Maryland 21701-5012

Contract No. DAMD17-87-C-7038

University of Alabama at Birmingham
701 19th Street South, LHRB 613
Birmingham, Alabama 35294

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188
Exp Date Jun 30, 1986

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; Distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION University of Alabama at Birmingham 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) 701 19th Street South, LHRB 613 Birmingham, AL 35294	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Med. Res. & Dev. Command 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAMD17-87-C-7038
8c. ADDRESS (City, State, and ZIP Code) Fort Detrick, Frederick, MD 21701-5012	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 623105 / PROJECT NO. 3M2 623105H29 / TASK NO. AD / WORK UNIT ACCESSION NO 045

11. TITLE (Include Security Classification)
(U) Studies of Genetic Variation in the AIDS Virus: Relevance to Disease Pathogenesis, Antiviral Therapy, and Vaccine Development

12. PERSONAL AUTHOR(S)
George M. Shaw, M.D., Ph.D.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 12/15/86 TO 12/14/89	1990 June 30	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	RA 1, AIDS, HTLV-III, Virology, HIV, Restriction Maps
06	03		
06	13		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Overall objectives for our contract studies (DAMD17-87-C-7038) were to define the extent and nature of HIV-1 and HIV-2 genetic variability in independent virus isolates from different geographic regions; the rate and characteristics of genetic change of HIV-1 *in vivo* in persistently infected individuals; and the biologic consequences of HIV variability in regard to cell tropism and virulence. The experimental approaches taken to address these important areas of viral pathogenesis included virus isolation, lambda phage cloning, nucleotide sequencing, transfection of proviral DNA, and biologic characterization of progeny virus. In addition to this, work was done to develop novel PCR and recombinant lambda phage cloning approaches that enabled the cloning and analysis of HIV/SIV viruses directly from uncultured lymphoid cells, thereby avoiding selective pressures inherent to tissue culture. The results of our studies are summarized below:

1. Envelope, gag and pol genes from (33) divergent HIV-1 isolates were cloned and sequenced. The envelope was shown to be hypervariable compared to other viral genes, to contain interspersed hypervariable and conserved domains, and yet to contain a conserved overall structure suggested by strong conservation of 18 cysteine residues. (Cell 45:637, 1986).

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Mary Frances Bostian	22b TELEPHONE (Include Area Code) 301-663-7325 / 22c OFFICE SYMBOL SGRD-RMI-S

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

19.

2. Sequential HIV-1 isolates from three individuals were obtained, cloned, and mapped, the envelope genes sequenced. Detailed molecular analysis showed the rate of evolution for the envelope gene *in vivo* to be $\geq 10^{-3}$ nucleotide substitutions per site per year, a rate 10^6 times that in eukaryotic genes and equal to that in influenza, Visna, and EIAV. Changes in the viral genomes consist primarily of isolated and clustered base substitutions, duplications, and short insertions or deletions suggesting that reverse transcriptase misincorporation and copy-choice misreading errors are responsible in large part for variation of HIV-1 *in vivo* (Science 232:1548, 1986).

3. Computer-assisted analysis of predicted hydrophilicity, surface probability, and flexibility identified putative antigenic sites in the viral envelope, generally in hypervariable regions but also in a conserved region as well (J. Virology 61:570, 1987).

4. Molecular analysis of sequential HIV-1 isolates from chronically-infected individuals demonstrated for the first time the "quasispecies" nature of HIV-1 *in vivo* and parallel evolution and persistence of complex mixtures of genotypically distinct viruses during persistent infection (Nature 334:440, 1988).

5. Isolation, cloning, and sequencing of virus from an infected lab worker was performed, providing evidence for down-regulation of viral replication *in vivo* yet progressive changes in the immunodominant envelope sequences (Science 239:68, 1988).

6. A biologically unique strain of HIV-2 (HIV-2/ST) was isolated from peripheral blood cells of a healthy Senegalese woman and shown to be attenuated in its ability to cause cell fusion and cell killing and in its ability to be transmitted cell-free. A full-length, replication competent proviral clone (λJSP4-27) was obtained from this isolate, shown to have the same phenotype as the parental isolate, and sequenced in its entirety (Science 240:1525, 1988; J. Virology 64:890, 1990). This clone, and a series of culture-derived revertants that were selected for cytopathic activity, now provide a substrate for structure-function analysis of the HIV-2 envelope gene and protein.

7. The molecular identity of "HTLV-4" was shown to be SIV_{MAC} (Nature 330:184, 1987). A novel reading frame termed X, present in SIV_{MAC} and HIV-2 but not HIV-1, was identified and shown to encode a 14kD protein that is packaged in the virion (J. Virology 62:3501, 1988).

8. A "nested" PCR amplification technique using generic or "universal" oligonucleotide primer pairs was developed that led to the detection, cloning, and sequencing of the first SIV_{AGM} virus strains from West African green monkeys. In addition, because of the increased sensitivity and specificity of the nested PCR technique, this could be done using uncultured peripheral blood mononuclear cell DNA without prior *in vitro* cultivation (manuscript in preparation).

9. Full-length, replication competent HIV-1 proviral clones were obtained by direct lambda phage cloning of DNA from uncultured human brain. These clones are the first such clones derived directly from human tissue without interim cell culture and correspond directly to the viral forms present *in vivo* in infected individuals. The progeny virus derived by transfection of COS-1 cells with this proviral DNA are highly replicative and monocyte tropic and are being characterized in regard to molecular determinants of monocyte tropism, biologic properties, genomic organization, gene structure-function relationships, and relevance to HIV-1 induced neuropathogenesis (manuscript in preparation).

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

GHJ Where copyrighted material is quoted, permission has been obtained to use such material.

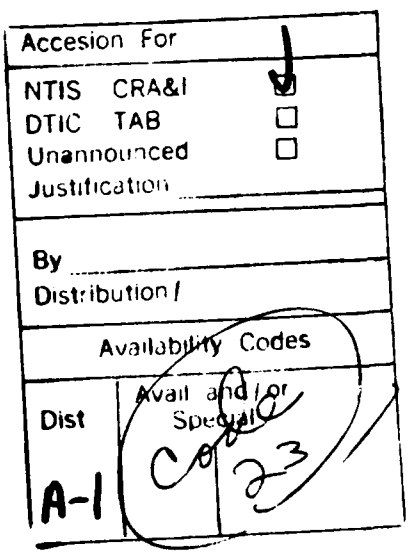
GHJ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

GHJ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

GHJ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

GHJ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

GHJ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institute of Health.



PI Signature Date

TABLE OF CONTENTS

When contract DAMD 17-87-C-7038 "Studies of Genetic Variation in the AIDS Virus: Relevance to Disease Pathogenesis, Antiviral Therapy, and Vaccine Development" was first initiated, relatively little was known about the genetics and biology of the HIV and SIV groups of viruses and their relation to disease pathogenesis (8,50). Over the relatively short period of time from 1981 to 1985, 15,000 cases of AIDS had been diagnosed in the United States alone, and it was estimated that an additional 40,000 cases would occur in the U.S. over the next two years (34,128). It was further estimated that up to 1-2 million Americans had already been infected with the HIV-1 virus, and it was not known what proportion of these would ultimately develop the full-blown AIDS syndrome. With a mortality rate approaching 100%, AIDS was recognized as clearly one of the most serious diseases confronting man in modern times. There were a number of reasons that AIDS and its causative agents, HIV-1 and HIV-2, were of prime importance to the U.S. Armed Forces and the civilian community alike. First, the AIDS virus causes a spectrum of serious and often fatal diseases and is transmitted through blood products and sexual relations. This meant that a great many persons, including military personnel who previously were not included in high-risk groups, were at considerable risk for infection. Second, the HIV is a "geographic agent" such that in certain locales including Africa, Haiti, and various U.S. and European cities its prevalence is particularly high. Third, while the AIDS epidemic affects persons of all ages it is particularly prevalent in young, sexually active adult men. Fourth, it is clear that HIV can be readily transmitted through heterosexual relations. Fifth, infection by HIV has a definite impact beyond clinically defined of AIDS. For example, for every one person with AIDS, there are 10 persons with an AIDS-related condition (ARC) such as lymphadenopathy, diarrhea, weight loss, hematologic abnormalities, or neurologic dysfunction and there are 100 more with asymptomatic infection.

Our laboratory had been, and continues to be, actively involved in molecular and genetic studies of HIV-1, HIV-2, and SIV. We reported the first successful cloning of HIV-1 (65,145), and in addition, determined that: (i) HIV-1 is an exogenous human retrovirus

approximately 10 kb in length that lacks nuclei acid sequences derived from normal human DNA; (ii) the virus persists in cells in both integrated and unintegrated forms; (iii) different isolates of the virus exhibit substantial diversity, or heterogeneity, in their restriction enzyme cleavage patterns; and (iv) HIV-1 frequently infects the brains of AIDS patients where it may lead to AIDS-associated dementia (65,66,69,143-145,164). The thrust of the proposed DoD work for contract DAMD 17-87-C-7038 was aimed at extending these studies. The reason for placing this emphasis on the genetic variation of HIV was that genetic variation, especially within the envelope gene, is likely critically involved in the virus's host-range and tissue specificity and in its ability to evade host defenses. It therefore follows that a thorough understanding of genetic variation will be essential for the elucidation of viral pathogenesis and for the development of effective vaccines, anti-viral agents, and diagnostic reagents for the HIV/SIV groups of viruses. Specific objectives of the project were:

1. To determine the extent, nature, and molecular basis of HIV-1 genomic variability.
2. To determine the rate and characteristics of HIV-1 genomic variation in chronically infected individuals over time.
3. To analyze the deduced amino acid sequence of the envelope of HIV-1 for predicted secondary structure and antigenicity.
4. To determine the molecular and biologic properties of HIV-2 isolates from individuals in West Africa where HIV-2 appears to be less virulent compared with HIV-1 and some other strains of HIV-2.
5. To determine the molecular basis for attenuation of cytopathicity in certain strains of HIV-2.
6. To develop PCR techniques for sensitive and specific amplification of novel HIV/SIV genomes using universal primers and "nested" amplification and to apply this to the detection and characterization of SIV and HIV-2 in uncultured blood

of West African green monkeys and people.

7. To molecularly clone full-length replication competent HIV-1 proviruses from uncultured human brain specimens for analysis of virus biology, molecular determinants of tropism and virulence, genomic organization, and gene structure-function relationships.

In this Final Report, work done to complete these scientific objectives is summarized.

1. **To determine the extent, nature, and molecular basis of HIV-1 genomic variability.**

In this initial phase of study, molecular clones of HIV-1 derived from virus cultures from blood samples of a Haitian homosexual man (HAT-3), a Haitian infant (WMJ), two U.S. homosexual men (ARV-2 and BH-10), and a French homosexual man (LAV-1) were studied (12,65,66,69,125,139). The genome of HIV-1 consists of long terminal repeat (LTR) elements; *gag, pol,* and *env* genes; as well as central region genes *vif, vpr, vpu, tat,* and *rev* and 3' orf for 3' open reading frame now known as *nef.* For HAT-3, we determined the full length proviral sequence. For WMJ-1, only the *env* gene was sequenced, and for two other WMJ clones that were closely related to WMJ-1, both *env* and *gag* (p15 and p24 regions) were sequenced. From this analysis, and from restriction enzyme mapping and heteroduplexing, we confirmed that the overall genomic organization of HAT-3 and WMJ-1 is the same as that described previously for BH-10, LAV-1, and ARV-2.

Figure 1 compares the nucleotide and deduced amino acid sequences of HAT-3, WMJ-1, LAV-1a, and ARV-2 to the original HTLV-IIIb clone, BH-10. These data demonstrate the existence of a broad spectrum of diversity among the AIDS retroviruses: Overall, BH-10 differed from HAT-3 in 8.6% of nucleotides (12.5% amino acids), from ARV-2 in 6.3% of nucleotides (9.2% amino acids), and from LAV-1a in 1.5% of nucleotides (2.2% amino acids). HAT-3 and ARV-2 differed from each other over the lengths of their sequenced genomes by 9.3% of nucleotides (14.2% amino acids). A similar spectrum of diversity among independent HIV-1 isolates has been reported on the basis of restriction enzyme mapping (13,164).

The distribution of sequence differences was not uniform throughout the viral genomes. Instead, when divergent genomes were compared, changes were found to be considerably more prevalent in the envelope and 3' orf (*nef*) than in other genes. In the envelope gene in particular, BH-10 differed from HAT-3 in 12.9% of nucleotides (19% amino acids), from ARV-2 in 9.8% of nucleotides (14.0% amino acids). In the same region, HAT-3 differed from ARV-2 in 14.7% of nucleotides (18.4% amino acids) and from WMJ-1 in 11.3% of nucleotides (16.2% amino acids). Within the envelope gene, the signal peptide and extracellular portion were the regions that were most highly divergent. For example, the extracellular envelope domains of HAT-3, WMJ-1, and ARV-2 differed from that of BH-10 in 21.4%, 18.7% and 17.0% of amino acids respectively. The two Haitian isolates, HAT-3 and WMJ-1, differed from each other in this region by 21.6% amino acids and from ARV-2 by 20.2% and 20.8%, respectively.

The most conserved areas in the genomes of HAT-3, LAV-1a, BH-10, and ARV-2 were the *gag* and *pol* genes, which differed among the viruses in less than 6% of nucleotides and less than 7% of amino acids. In these regions, nucleotide sequence changes were almost exclusively due to point mutations in contrast to *env* where clustered nucleotide changes involving in-frame deletions, insertions, and/or duplications were common. For example, compared to BH-10, there were 48bp of insertions and 21 bp of deletions in the extracellular envelope of HAT-3 as compared to only 6 bp of insertions and deletions in all of its analyzed *gag, pol,* and *vif* sequences. Similarly, there were 33 bp of insertions and 36 bp of deletions in the extracellular envelope of ARV-2 (compared to BH-10) but only a 6 bp insertion and no deletions in al of its *gag, pol,* and sor (*vif*). Even in LAV-1a, which is more closely related to BH-10 than are the other sequenced viruses, there was a 15 bp insertion (duplication) in the exterior envelope gene but no insertions or deletions in *gag, pol,* or sor (*vif*).

Another difference in the types of mutations present in the *env* gene compared to the

gag gene of these viruses was in the proportion of silent third base pair changes. Excluding deletions and insertions, compared to BH-10 the mutations in the external *env* genes of ARV-2, HAT-3, and WMJ (clone WMJ-2 from patient WMJ) were in the third position of the codon in 36%, 37% and 34% of instances leading to 32%, 25% and 42% amino acid changes, respectively. Conversely, in the *gag* genes of these same viruses, 63%, 56%, and 66% of nucleotide changes were in the third place of the codon leading to 10%, 3.5% and 0% amino acid changes, respectively. Thus, more than half of the single nucleotide changes in *env* occurred in the first or second codon position, and even for third position changes, many of these led to amino acid changes. Conversely, only a minority of first or second codon position changes occurred in *gag*, and in this gene the third position changes were almost uniformly silent. These findings are consistent with two interpretations; first, that structure-function relationships lead to stronger conservation of amino acid sequence in *gag* than in *env*, and second, that non-silent nucleotide changes in *env* may actually be under positive selection pressure. Whether or not the latter interpretation is correct must await a much more complete understanding of the host immunologic response to HIV-1.

An alignment of the envelope nucleotide and amino acid sequences of HAT-3, WMJ-1 BH-10, LAV-1a, and ARV-2 is shown in Figures 2 and 3. A methionine codon at position 8 of the BH-10 envelope open reading frame most likely marks the beginning of the gene, with a hydrophobic region between position 17 and 37 representing the potential leader sequence, or signal peptide. This signal peptide is cleaved from the envelope precursor protein during envelope maturation (5). The ensuing peptide sequence from position 38 to 518 (BH-10 sequence) is slightly hydrophilic and contains numerous potential N-linked glycosylation sites, ranging from 22 in WMJ-1 to 29 in HAT-3. This region corresponds to the major exterior envelope glycoprotein (gp120). A characteristic arginine-rich hydrophobic stretch marks the cleavage site for the processing of the envelope precursor gp160 into the exterior gp120 and membrane-bound gp41 (5,42). The latter protein includes an apparent hydrophobic membrane-

spanning segment, a hydrophilic anchor sequence, and an additional carboxy-terminal stretch of 150 residues whose function is presently unknown.

A striking feature of the five envelope sequences was the conservation of cysteine residues (Figure 3). Within the extracellular envelope, each of the 18 cysteine residues was conserved in all five viruses. Within the transmembrane protein, all but one cysteine residue in its extreme carboxy-terminus was conserved. This finding argues for a highly conserved "macrostructure" of different HIV-1 envelope glycoproteins. Despite this overall conservative structure numerous amino acid changes were evident throughout the envelope glycoprotein. In the transmembrane region these were generally the result of isolated nucleotide point mutations leading to single amino acid substitutions. In the extracellular region, changes in amino acid sequence resulted from both nucleotide point mutations and in-frame deletions, insertions or duplications of blocks of nucleotides. The resulting amino acid alterations appeared as clustered mutations interspersed with polypeptide segments that were highly conserved. The regions of highest variability (dark shading in Figure 3) and highest conservation (light shading in Figure 3) were identified both by visual inspection and by objective determination of relative variation using computer analysis (Figure 4).

Since the extracellular region of the HIV-1 envelope glycoprotein constitutes a major target of the host immunologic response (10), we examined this region for predicted antigenic epitopes using a computer program that predicts the secondary structure of proteins superimposed with values for hydrophilicity (24,80). Such an analysis of other proteins, including viral envelopes has shown that antigenic epitopes are often associated with hydrophilic protein domains containing β turns (7,32,46,62,121,160). This analysis demonstrated that the exterior envelope proteins of the five AIDS retroviruses each contain a number of sites that meet criteria for likely antigenic epitopes and that these regions generally coincide with the variable regions identified independently by amino acid sequence comparisons. Figure 5 illustrates the substantial differences in predicted secondary protein

structure for the first variable region of the external envelope gene product for five viruses. In this region, there were 7 β turns in HAT-3, 0 β turns in WMJ-1, 8 β turns in BH-10, 11 β turns in LAV-1a, and 5 β turns in ARV-2. The five envelope sequences also differed in this region in their degrees of hydrophilicity and in the number and location of potential N-linked glycosylation sites (Figures 2 and 3). In each of the other variable regions substantial differences existed similarly in predicted hydrophilicity, secondary structure, and potential glycosylation sites.

Interspersed with the variable regions of the exterior envelope protein were other areas that were highly conserved among all five isolates analyzed. Most of these regions were primarily hydrophobic and contained only few or no β turns. Accordingly, they are believed to be less likely to represent antigenic epitopes. There were, however, a number of exceptions. One such exception is a conserved stretch of 46 amino acid residues immediately adjacent to the processing site of the envelope precursor (amino acids 473 through 518 in Figure 3). This area was very highly conserved among all five viruses, contained numerous β turns, and was hydrophilic. Thus this conserved region of the exterior envelope glycoprotein would be expected to be both antigenic and cross-reactive among different viral strains.

2. To determine the rate and characteristics of HIV-1 genomic variation in chronically infected individuals over time.

The results shown in Figures 1-5 indicated that (i) within the extracellular *env* gene there are localized regions of hypervariability interspersed with regions of strong conservation; (ii) genetic changes among different viruses result largely from duplications, insertions, or deletions of short stretches of nucleotides as well as from an accumulation of nucleotide point mutations; and (iii) hypervariable *env* regions, and some constant regions, have properties predictive of antigenicity based on analysis of predicted secondary protein structure, hydrophilicity, and glycosylation pattern. To define further the extent, rate and nature of genetic variation of HIV-1 *in vivo* we have studied sequential virus isolates from three

persistently infected patients with AIDS or risk factors for AIDS. The isolates were obtained on four to six occasions over a 1- or 2-year period. The first patient (coded WMJ) was a Haitian child with AIDS who was born in Miami, Florida, in 1982 and infected perinatally by her HIV-1-positive mother. The second patient (coded WMF) was a 34-year-old Haitian woman with AIDS-related complex (ARC) who was also from Miami and whose only known risk factor for HIV-1 infection was heterosexual exposure. The third patient (coded RJS) was a 31-year-old American homosexual man from California who, despite chronic HIV-1 infection, remained clinically well for 5 years but eventually developed evidence of T-cell dysfunction and died of AIDS. He reported having sexual encounters with at least 1000 different men between 1980 and 1985. These three subjects thus represented a broad clinical spectrum of HIV-1-related disease (AIDS, ARC, and clinically normal virus carrier), different modes of virus transmission (perinatal, heterosexual, and homosexual), and widely different numbers of exposures to the AIDS virus (presumably one for WMJ, limited encounters for WMF, and very large numbers for RJS). All virus isolations were made from the patients' peripheral blood mononuclear cells (PBMC) that were cocultivated with normal PBMC and, in most instances, transmitted to immortalized T cell lines.

Figure 6, shows that, for patient WMJ the Sst I, EcoRI, and Pst I restriction enzyme patterns for each of the five isolates were identical and that the Bgl II and Pvu II patterns were almost identical. With Hind III, the isolates were clearly distinct but still highly related; for example, the loss of a single Hind III site in WMJ-1 led to the Hind III pattern observed in WMJ-2 (See Figure 7). In contrast, the enzyme cleavage patterns of the five WMJ isolates, which as a group were quite similar to each other, differed markedly from viruses of patients WMF and RJS as well as from all other patients' viruses tested. The observation that serial virus isolates from a single patient (WMJ) were more highly related to each other than to viruses from other patients was also true for isolates from WMF and RJS. For example, the four isolates from WMF were identical in Sst I and EcoRI patterns and nearly identical when

analyzed with Pvu II and Pst I. Moreover, the differences with Hind III could be explained simply by either the loss or gain of single restriction sites. For RJS, all six isolates were identical with Sst I and Eco RI and nearly identical with Bgl II, Pvu II, and Pst I.

For most virus isolates shown in Figure 6, the sizes of the restriction fragments generated by any given enzyme added up to 9 kilobases, the approximate genomic size of HIV-1. However, for certain virus isolates from patients WMF and RJS, some of the restriction enzymes generated fragments that added up to more than one genomic equivalent (for example, Bgl II, Pvu II, and Pst I digestions of RJS-4; the Hind III patterns of RJS-2, RJS-5, and RJS-6; the Bgl II of WMF-1; and the Pst I pattern of WMF-3). These particular isolates were therefore composed of a mixture of more than one highly related, yet genotypically distinct, virus populations. That the restriction digestions did, in fact, reflect accurately the presence of more than one predominant viral genotype and were not reflecting incomplete digestions was shown by the following: (i) several different enzymes generated bands whose additive size in each case was greater than 9 kb; (ii) repeated digestions of these same DNA's with excess enzyme gave identical result; and (iii) molecular clones of viral DNA obtained from several of these DNA's corresponded to the different viral patterns apparent on the blot hybridizations.

Although more than one predominant viral genotype was present in some virus isolates, all of the genotypically distinct viral forms in any one patient were still very similar to each other as opposed to independent virus isolates from different individuals that were distinct. For example, although two viral forms could be distinguished by the Bgl II digestion of WMF-1, these two viral forms were identical in their Sst I, Hind III, EcoRI, Pvu II, and Pst I patterns. Similarly, the two viral forms distinguishable by Hind III in RJS-2, RJS-5, and RJS-6 were identical in Sst I, EcoRI, Pvu II, and Pst I patterns. By way of contrast, the WMF viruses and RJS viruses, as groups, differed from each other dramatically in Sst I, Hind III, Bgl II, Pvu II, and Pst I patterns. Thus, within any given individual, even one presumably exposed to

numerous different AIDS viruses, only one or a very limited number of virus strains, as defined by restriction enzyme analysis, could be detected by virus isolation and DNA blot hybridization. It is possible, of course, that these patients were actually infected with many different genotypic viral forms but, because of their unequal growth characteristics, most were not evident after cultivation *in vitro*. We cannot rule out this possibility, but it is of note that work involving Southern blot hybridization analyses of uncultured brain and lymphoid tissues from AIDS patients also indicates the presence of only one or a very limited number of predominant viral forms *in situ* (135,144).

To define in greater detail the nature of genetic changes in HIV-1 that occur in infected patients over time, we used subgenomic probes to map the restriction enzyme cleavage patterns of isolates WMJ-1 through WMJ-5 (Figure 7). In addition, the proviral genomes of WMJ-1, WMJ-2, and WMJ-3 were molecularly cloned and the nucleotide and deduced amino acid sequences of their *env* genes determined (Figure 8). The restriction maps in Figure 7 represent the restriction cleavage patterns of proviral DNA for each of the five WMJ isolates shown in Figure 6. The results indicated that each of the five WMJ isolates that were obtained over a 1-year period were highly related to each other, differing in only one or a few restriction sites. For example, out of 29 restriction sites tested, WMJ-2 differed from WMJ-1 in only four sites and from WMJ-3, WMJ-4, and WMJ-5 in three, one, and one sites, respectively. Of particular note was the finding that although the five WMJ virus isolates were obtained sequentially, these viruses had apparently not evolved genetically in such a direct fashion, that is, from WMJ-1 to WMJ-2 to WMJ-3 and so on. This could be deduced from the restriction maps by observing the nondirectional changes in certain of the Pvu II, Hind III, Kpn I, and Bgl II sites (Figure 7).

That isolates WMJ-1, WMJ-2, WMJ-3, WMJ-4, and WMJ-5 had evolved in parallel was confirmed by determining and comparing the *env* nucleotide sequences of molecular clones of proviral DNA from WMJ-1, WMJ-2, and WMJ-3 (Figure 8). Overall, the *env* of WMJ-1

(clone λWMJ-1-V) differed from WMJ-2 (clone λWMJ-2-III) in 2.8% of nucleotides and that of WMJ-2 differed from WMJ-3 (clone λWMJ-3-III) in 3.1% of nucleotides. If WMJ-3 had evolved directly from WMJ-2, which had in turn evolved from WMJ-1, one would have expected to find approximately 5% to 6% nucleotide differences between WMJ-3 and WMJ-1. This was not the case because WMJ-3 differed from WMJ-1 by only 2.3% of nucleotides. Direct inspection of the nucleotide alignment of WMJ-1, WMJ-2 and WMJ-3 (Figure 8) also revealed many examples where WMJ-1 was more similar to WMJ-3 than to WMJ-2, and vice versa. Of importance, the restriction patterns of each of the three WMJ proviral clones that were sequenced corresponded exactly (20 out of 20 restriction sites) to the predominant genomic proviral forms for WMJ-1, -2, and -3 present in infected cells and shown in Figures 6 and 7. Furthermore, between 5 and 20 proviral DNA clones were obtained from the recombinant libraries or each isolate, and again, the ones chosen for sequence analysis were full-length and representative of the majority. Finally, to be certain that the observed nucleotide differences among the clones were not an artifact due to technical procedures, we sequenced a portion of the *env* gene of two proviral clones from the same library. These two clones (λWMJ-2-III and λWMJ-2-XIII), both isolated from a recombinant λ phage library of WMJ-2 DNA, were identical in 1220 out of 1221 nucleotides sequenced. Thus, the sequenced proviral clones were representative of the predominant viral forms in each of the isolates, and observed sequence differences reflected variation in the HIV-1 *env* gene.

Since it was shown that WMJ-1, WMJ-2, and WMJ-3 were highly related and had evolved from a common progenitor, and since the date of infection of patient WMJ by HIV-1 was known, it was possible to determine the approximate rate of genetic evolution for the AIDS virus during a natural infection. Many data documenting the rates of evolution of other RNA viruses as well as DNA genomes had led to the conclusion that RNA viruses in general evolved at a rate much greater than DNA genomes (43,79). For calculating the rate of genetic change for HIV-1, we assumed that WMJ-1, WMJ-2, and WMJ-3 had all evolved from a

common progenitor virus sometime within the preceding 5 years. This supposition was based on the following information (i) Patient WMJ, who had only a single perinatal exposure to HIV-1, was 2 years old when isolates 1, 2, and 3 were obtained. (ii) WMJ-1, -2, and -3 were but three of five highly related viruses isolated from this same patient. (iii) All five WMJ viruses were highly related to each other and very different, as a group, from other independent viruses that we had evaluated (including five other isolates from Haitian individuals living in Miami). (iv) Within the hypervariable envelope regions, the nucleotide sequences of WMJ-1, -2, and -3 were very similar to each other and dramatically different from those of isolates from other patients. These data, taken together, indicate that the five WMJ isolates all evolved from a common progenitor virus. Conceivably, patient WMJ could have been infected perinatally with more than one virus. Even if this were so, however, these viruses would still have had to evolve in the immediate past from a common progenitor virus present in WMJ's mother.

Using 1 to 5 years as an estimate of the time since divergence of WMJ-1, WMJ-2, and WMJ-3 from a common point, we calculated the rates of evolution for the *env* and *gag* genes of HIV-1 as follows:

$$R = D/2T$$

where T is the rate of nucleotide substitutions per site per year and T is the time since divergence of the viruses

$$D = -(3/4) \ln [1 - (4/3)P]$$

where P is the proportion of different nucleotides between the homologous genes. From this equation and the determined nucleotide sequences of WMJ-2 and WMJ-3, the rate of evolution (R) for the HIV-1 *env* gene was estimated to be between 1.58×10^{-2} and 3.17×10^{-3} nucleotide substitutions per site per year for T values of 1 year and 5 years, respectively. Similarly, the rate of genetic mutation for the *gag* p24 and p15 genes was estimated to be between 1.85×10^{-3} and 3.70×10^{-4}. This rate of genetic change for the AIDS virus is more

than a millionfold greater than for most DNA genomes and may even be tenfold greater than from some other RNA viruses including certain retroviruses and influenza A virus.

We also compared the nature and distribution of nucleotide changes within the *env* genes of sequentially isolated viruses from persistently infected individuals with those in independent virus isolates from different patients. The changes in the *env* genes of WMJ-1, WMJ-2, and WMJ-3 were almost exclusively nucleotide point mutations. This corroborates our earlier findings with independent virus isolates that single and clustered base pair substitutions, presumably resulting from errors in transcription, are an important source of variation of HIV-1. In addition to point mutations, however, there were also three small (3 bp) deletions in WMJ-2 and WMJ-3 compared to WMJ-1 (figure 3). Their presence underscores the importance of nucleotide deletions or insertions as well as substitutions as a mechanisms for genomic change in the AIDS virus, a conclusion also supported by analysis of independent HIV-1 isolates. The distribution of nucleotide substitutions and deletions or insertions in the three WMJ isolates was not uniform throughout the *env* gene. Instead, such changes were clustered, coinciding with hypervariable regions previously identified in five independent AIDS virus isolates (see Figure 9).

In the studies of HIV-1 variation described in Section 1, we observed that independent as well as sequential virus isolates from individual patients showed evidence of rapid genetic change. Genotypic analysis of paired HIV-1 isolates from donor-recipient transfusion cases (148) and nucleotide sequence comparisons of geographically distinct HIV-1 strains (3) provided further evidence for the highly variable nature of HIV-1. However, the spectrum of HIV-1 variation in chronically infected individuals, the mechanisms underlying this variation, and its biological consequences still remained unknown. The experiments described in Figures 6-9 provide insights into this area. To address this question further, we analyzed a group of molecular proviral clones from sequential virus isolates from two infected individuals (RJS and WMF) with HIV-1 associated clinical disease. Recombinant lambda phage libraries of one RJS

isolate (RJS4) and two WMF isolates (WMF1 and WMF3) were prepared so that the viral DNA molecules which *in toto* comprise the hybridization patterns of the overall isolate DNA (Figure 6) could be dissected and analyzed individually. A total of 27, 17, and 18 full-length HIV-1 clones were obtained from the three libraries (Figure 10). Careful restriction enzyme mapping using 11 endonucleases revealed that 17 of the 27 RJS4 clones, 10 of the 17 WMF1 clones, and 13 of the 18 WMF3 clones were distinguishable by unique cleavage patterns. The remarkably large number of distinguishable viral clones within each isolate resulted from various combinations of restriction site polymorphisms distributed throughout the viral genomes. As expected, the restriction patterns of the predominant HIV-1 clones corresponded with the fragments visualized on blot-hybridizations of the isolate DNA from which they were cloned. Because each restriction enzyme generated only one to three different fragmentation patterns, the many unique viral genotypes shown in this report to comprise a virus isolate had previously gone unrecognized.

Inspection of the genomic restriction patterns of the HIV-1 clones in Figure 10 showed that clones from individual virus isolates (RJS4a-q; WMF1a-j; WMF1a-m) were considerably more similar to other clones within that isolate than to unrelated clones such as HXB2, LAV-Mal, LAV-Eli, ARV2, and WMJ1. In order to quantify objectively the extent of similarity among the different viral clones, a pairwise comparison of the percentage of restriction site differences between each of the 45 viral genotypes shown in Figure 10 was performed (total of 1,035 independent comparisons; Figure 11). Such an analysis of restriction site differences between related clones is a valid means for estimating overall nucleotide sequence variability (40,91). Assuming that the loss of a restriction site in otherwise highly related genomes results from a change in a single nucleotide, a 50% difference in restriction sites when using enzymes that recognize 6 base pair sequences corresponds to approximately 8% nucleotide sequence difference (i.e., 1 nucleotide change out of every 12 nucleotides sampled). We used this approach first to analyze clones of HIV-1 for which nucleotide sequence information was

available, and the correlation between nucleotide sequence differences and restriction site differences was found to be quite good. For example, clone HXB2 differs from clones BH10, ARV2, and LAV-Eli by 13%, 41%, and 62% in restriction sites (ref. 145 and Figure 11). Based on these restriction site differences, the predicted nucleotide differences would be 2.2%, 6.8%, and 10.5%. The actual differences determined by nucleotide sequence comparisons are 1.6%, 5.8%, and 9.7%, respectively (59). Similarly, clones of WMJ2 and WMJ3, both of which were derived from a single HIV-1 infected patient, differ by 12% in restriction sites and by approximately 1-2% in nucleotide sequence (69). Clones of LAV-Mal, LAV-Eli, and ARV2 differ from each other by 58 - 70% in restriction sites (Figure 11) and by 10.1 to 13.0% in nucleotide sequence. In the present study, the restriction cleavage site differences among all 45 clones ranged from 3% to 70%. The viral DNA genomes comprising isolates RJS4, WMF1 and WMF3 were considerably more similar to other viral genomes from within the same isolate than to viral genomes from unrelated (independent) isolates. For example, the 17 different RJS4 genotypes varied from each other by 3 - 28% (mean of 13%) whereas the same clones varied from independent viruses by 41 - 70% (mean of 55%; $p<0.0001$). The 10 different WMF1 genotypes varied from each other by 3 - 33% (mean of 13%) and from independent viruses by 34 - 64% (mean of 54%; $p<0.0001$). The 13 different WMF3 genotypes varied from each other by 4 - 29% (mean of 13%) and from independent viruses by 38 - 65% (mean of 56%; $p<0.0001$). These data show that while considerable genotypic diversity of viruses exists within a given isolate, these multiple distinguishable viral forms have all clearly evolved from one another, or from common precursor viruses, and do not represent concomitant infection by unrelated (independent) viruses.

Isolates WMF1 and WMF3 were derived from cultures of peripheral blood lymphocytes of the same individual taken 16 months apart. If the viral forms present in WMF3 had evolved sequentially from, or in parallel with, the viruses present in WMF1, the extent of similarity between WMF1 and WMF3 viral genomes should be intermediate between values for clones

from within each isolate and clones from independent (unrelated) isolates. If superinfection with unrelated viruses had occurred, genotypes of WMF1 and WMF3 viruses would be expected to vary to the same extent as the unrelated viruses HXB2, LAV-Mal, LAV-Eli, ARV2, WMJ1, RJS4a, and WMF1a. The data in Figure 11 show that the former, not the latter, was the case. The viral clones WMF1 (a-j) differed from clones WMF3 (a-m) by 11 - 44% (mean of 23%), an intermediate value statistically different from those for viruses from within single isolates (mean of 13%; $p < 0.0001$) or unrelated (independent) isolates (mean of 55%; $p < 0.0001$). These data, plus the unique similarities among WMF1 and WMF3 clones depicted in Figure 10, indicate that the viruses comprising the latter isolate evolved either from genomes present in the former isolate or from immediate precursor genomes common to both. The absence of clones in WMF3 that are identical to clones in WMF1, and the existence of certain viral genotypes in WMF1 and WMF3 that differ by as much as 39% in restriction sites (e.g., WMF1h and WMF3m), underscores the extremely rapid rate of change of HIV-1 *in vivo*.

As a control for these studies, and to determine the extent to which *in vitro* changes in HIV-1 could contribute to the observed genetic differences in RJS and WMF isolates, we determined the extent of HIV-1 genotypic variation occurring during cultivation and amplification of virus from infected patient tissues. Peripheral blood mononuclear cells (PBMC) obtained from patient RJS at the same time that isolate RJS4 had been obtained were divided into two parts and each was co-cultivated with PHA-stimulated PBMC from different HLA-unrelated normal donors. The restriction cleavage patterns of the two virus isolates obtained after 6 weeks of culture were identical (Figure 12A), implying that the virus populations isolated were representative of viruses present *in vivo*. PBMC taken from the same patient two and four weeks after the first phlebotomy yielded virus populations that were genetically related to the first isolates yet distinguishable (Figure 12B), reflecting the recognized genotypic heterogeneity of HIV-1 *in vivo*. Finally, the genotypic pattern of virus cultured from the brain of an HIV-1 infected patient (BB) was compared to viral DNA extracted directly from an *uncultured*

specimen of the same tissue. The restriction patterns were identical (Figure 12C) indicating that the overall genotypic pattern of virus isolated by short-term culture corresponds to virus populations actually present *in vivo*. In a separate set of experiments, a recombinant lambda phage library identical to those prepared for RJS4, WMF1, and WMF3 was prepared from an isolate of HIV-1 that had been biologically-cloned by five sequential cell-free end-point dilutions of culture supernatant. This was done to obtain a genotypically-pure virus stock that could then be expanded in culture, as is done during the isolation of HIV-1 from human tissues, in order to determine how much genotypic variation is introduced during virus amplification in vitro. From this library, 10 full-length viral DNA clones were identified and these were identical in 270 out of 273 restriction sites mapped (average of 27 sites per clone; data not shown). This degree of restriction enzyme site variability (1%) in a biologically-cloned isolate that was amplified through multiple rounds of replication was significantly less than for the isolates RJS4 (13%), WPF1 (13%), and WPF3 (13%) (Figure 11). These data thus suggest that the extensive genotypic variation identified within naturally-occurring isolates of HIV-1 described in the present study are not the result of changes occurring during *in vitro* virus propagation or molecular cloning, although it is recognized that under conditions of selective pressure significant genetic changes in the HIV-1 genome can and do occur *in vitro* (162).

The results of this study indicate that genotypic variation of HIV-1 *in vivo* is rapid and extensive, that numerous variant viral forms coexist over time within the same patient, and that "isolates" of HIV-1 actually consist of complex mixtures of genotypically-distinct, albeit related, viruses. The present study also indicates that during natural infection by HIV-1, different viral genomes evolve in *parallel* and result in the emergence and persistence of multiple distinct genotypic forms.

Variation of HIV-1 has parallels in other lentiviral systems including equine infectious anemia virus (EIAV), visna virus, and simian immunodeficiency virus (SIV) (119,120,154). For EIAV, it is clear that genotypic changes are responsible for biologically important alterations

in viral antigenicity which allow the virus to elude host immune defenses (113). For visna, the significance of antigenic variation is less clear (154). In the SIV system, a virus strain has recently been isolated from a pig-tailed macaque that possesses altered biologic and antigenic properties leading to a broader host-range and a rapid, fatal immunodeficiency syndrome several weeks after inoculation. There are indications that genotypic variation in HIV-1 is similarly associated with potentially important biologic differences among variant forms. We have examined the biologic properties of viruses derived by transfection of full-length WMF1.16 and WMF3.3 clones (patterns WMF1a and WMF3c in Figure 10) and found them to be indistinguishable from wild-type virus in T-cell tropism and cytopathicity (data not shown). We have also examined the biologic properties of hybrid viruses constructed by exchanging the envelope region of six different RJS4 clones (numbers 6, 15, 16, 22, 24, and 26) into the transfection-competent HIV-1 prototype clone HXB-2 (51) and have found that the progeny viruses exhibit striking biological differences. These findings are especially intriguing in light of recent findings by Mullins and co-workers who showed that the envelope/LTR region of a replication-defective variant of feline leukemia virus (FeLV), when introduced into a replication competent construct of FeLV, was responsible for inducing a highly reproducible fatal immunodeficiency illness in cats (118). Furthermore, Gartner and Popovic have shown that some isolates of HIV-1 preferentially replicate in mononuclear phagocytes whereas others show preference for T-lymphocytes (58), and Koyanagi and Chen have identified genotypically-distinct virus that replicate in mononuclear phagocytes and brain glioma explant cultures (96). Our results demonstrate the extreme plasticity of the HIV-1 genome as a genetic substrate for important biological variation. Finally, the observation that HIV-1 "isolates" generally consist of complex mixtures of genotypically-distinct viruses is in keeping with the general concept of RNA viruses as "quasispecies" (43,150). Experiments examining the genetic, biologic, and immunologic properties of HIV-1 will need to account for this extensive genotypic heterogeneity, and in some instances may require the use of biologically or molecularly cloned viruses.

3. To analyze the deduced amino acid sequence of the envelope of HIV-1 for predicted secondary structure and antigenicity.

Most patients with acquired immunodeficiency syndrome (AIDS) or AIDS-related complex show specific antibodies directed against proteins of human immunodeficiency virus (HIV), which have virus neutralizing activity (10,133,159) and are supposed to be directed against antigenic determinants located on the surface glycoprotein, as has been shown for other enveloped virus particles. However, the virus seems to have adopted properties which allow evasion of the immune surveillance mechanisms of the host. Differences among various isolates of HIV have primarily been analyzed at the nucleotide sequence level of independent isolates and also in sequential viral isolates from the same patient. These variations seem to be concentrated in the envelope protein-encoding region of HIV and may be fundamentally important for the biology and pathogenicity of HIV. For this reason, and for the development of viral antigens for diagnostic or vaccine use, it is important to identify and characterize antigenic determinants located in the glycoprotein complex of HIV and to define their possible functions.

We analyzed the amino acid sequences of the envelope protein complexes derived from the nucleotide sequences of seven AIDS virus isolates, three of which represented sequential isolates from the same patient. This work was an extension of the work described above in sections 1 and 2 on the genetic variability of the HIV envelope protein complex which mainly focused on the DNA and primary amino acid sequences. By computer analysis we predicted the secondary structure of gp120 and gp41, the cleavage products of gp160, and predicted potential antigenic sites by superimposing this secondary structure with the values for hydrophilicity, flexibility, surface probability, and glycosylation. Eleven potential antigenic sites were identified, 9 of which were located in the exterior part (gp120) of the envelope protein and 2 in the membrane-bound portion (gp41). Five highly variable regions were characterized, all contained in gp120 coinciding with the predicted epitopes. In sequential isolates from a

single patient, all alterations of secondary structures occurred in those regions which were identified as probable antigenic epitopes.

The open reading frames of the envelope gene derived from the nucleotide sequence of the seven virus isolates are similar in size and encode 854 (WMJ3) to 873 (HAT3) amino acids (Figure 13). Methionine codons at positions 8 (BH10 and LAV1A) or 9 (HAT3, ARV2, WMJ1, WMJ2, and WMJ3) mark the beginning of the envelope protein with a potential leader sequence, which is cleaved from the envelope precursor protein during maturation. The peptide sequences from position 38 represent the envelope precursor gp160, which is cleaved into the exterior gp120 (with 23 to 25 potential N-glycosylation sites) and the membrane-bound gp41 (containing 4 to 7 potential N-glycosylation sites). A stretch of positively charged amino acids at positions 512, 509, 508 (WMJ1, WMJ2, WMJ3), 517 (ARV2), 518 (BH10), 523 (LAV1A), and 275 (HAT3), respectively, marks the cleavage site (5,42).

Although the overall sizes and structures of the seven surface proteins are rather similar, the deduced amino acid sequences differ substantially. On the average only 66% of the amino acids are conserved in the exterior part of the protein, and these changes are clustered in special regions with only up to 10% conserved amino acid. gp41, the transmembrane part of the envelope protein complex, shows more than 80% conserved amino acids (Figure 13) and no regions of high variability. Furthermore, the changes in the latter region are all due to point mutations, whereas changes in gp120 frequently result from insertions and deletions and appear as clustered mutations interspersed with segments which have a high content of conserved amino acids (89%).

According to their content of conserved amino acids, glycosylation sites, and β-turn regions, the envelope proteins of HIV were subdivided into highly variable and constant regions. Constant regions were defined to contain 75% or more conserved amino acids and to have no amino acid insertions and deletions. In contrast, variable regions showed a low degree of conserved residues (25% or less) and a great variability in length due to deletions

or insertions; these changes occurred at least every fifth amino acid.

Due to the high number of conserved amino acid residues in the constant regions, glycosylation sites and secondary structures also showed a low degree of variation; more than 50% of potential glycosylation sites and amino acids in β-turn configurations were found to be conserved. Regions of high variability had a low degree of conserved β-turns (0 to 25%), especially in a hydrophilic environment, and almost no conserved potential glycosylation sites (0 to 25%). By these criteria, surface glycoprotein gp160 could be subdivided into a clustered pattern of highly variable (V1 to V5) and constant regions (C1 to C6). For purposes of classification the minimum length of a region was set at 10 amino acid residues; the locations and parameters of those regions are shown in Figures 13-16.

For further analysis we predicted antigenic determinants in the amino acid sequences with a computer program which predicted the secondary structure and calculated the values for hydrophilicity, flexibility, and surface probability (Figures 14 and 15). These regions are mainly located in β-turn regions which show a high degree of nonhydrophobic of flexible amino acid sequences or are predicted to have a high probability of location at the surface of the polypeptide (32,46,47,62,121).

In gp120, nine epitopes (I to IX) with a high antigenic potential can be predicted (Figure 16). Epitope I is located in a region of high variability (V1) at amino acids 137 to 154 of viral strain WMJ1. The locations of the antigenic epitopes for the other isolates are indicated in Table 2.

Epitope I has elevated values for hydrophilicity and flexibility, 1 to 3 potential glycosylation sites, and a number (up to 11 in LAV-1a) of amino acids in β turn configurations. Isolates WMJ1 and WMJ3 show no β turns due to the high variability of this region. Epitope II (amino acids 186 to 203) is situated in variable region V2 and contain s residues with high values for hydrophilicity, flexibility, and surface probability; up to 15 amino acids (HAT-3) have β turn configurations (with the exception of ARV2), and all strains may be glycosylated.

Epitope III (amino acids 232 to 246) in constant region C2 shows two conserved hydrophilic β turns, two conserved potential glycosylation sites, and high values of flexibility. Epitope IV (Amino acids 300 to 320) in variable region V3 shows 2 to 8 amino acids in β turns, with elevated values for hydrophilicity, flexibility, and surface probability, and has 1 to 2 potential glycosylation sites. Epitope V (amino acids 358 to 375) is only slightly variable (72% conserved amino acids), has high values for hydrophilicity, surface probability and flexibility, and has 1 to 2 N-glycosylation sites and 6 to 8 residues in β turn configurations. Epitope VI (amino acids 394 to 412), in contrast, is in highly variable region V4, with 2 to 4 possible glycosylation sites, 4 to 11 β turns, and high values for hydrophilicity and flexibility; the values for surface probability are not elevated. Epitope VII (amino acids 445 to 458) in constant region C3 has 4 to 7 β turns. One glycosylation site is hydrophilic and flexible and is directly followed by epitope VIII (amino acids 459 to 469), which shows high variability (V%), is hydrophilic, flexible, and possibly glycosylated, and has one (ARV2) to six residues in β turn configurations. Epitope IX (amino acids 470-483) in region C4 has high values for all parameters, four conserved β turns, and no potential glycosylation sites.

Only two antigenic sites could be identified in gp41. These contain amino acids 612 to 613 and 722 to 745 (Figures 13-16). The first epitope X, is located in a slightly variable region (56% conserved amino acids) and contains four conserved glycosylation sites, has high values for hydrophilicity, surface probability, and flexibility, has 4 to 6 β turns, and is probably the only antigenic site in gp41 which is located outside the lipid bilayer and accessible to antibody reaction and recognition. The second epitope (amino acids 722 to 745) is located directly after a stretch of hydrophobic amino acids which is likely to be transmembrane region of gp41 (TM3); these hydrophilic, flexible amino acids in β turn regions with high surface probability probably represent the hydrophilic anchor sequence which has been identified in most membrane proteins. This region, however, should be inside the cell and thus is not the best epitope for antigenic response. The following stretch of about 100 amino acids has a

further region which might be an antigenic determinant. This region, whose function is not known, might however, be gradually cleaved off by proteases from the precursor before maturation of gp41. With antibodies against a synthetic peptide derived from this region, mainly the precursor gp160 could be identified; furthermore, the molecular size of a glycosylated gp41 should be about 52,000 to 54,000 daltons by calculation (42,000 daltons of the primary product plus the molecular size of the carbohydrate moiety) unless it is proteolytically processed.

The envelope proteins of the sequential viral isolates from the same patient which were taken at intervals of 3 and 4 months are very similar in size and show only three amino acid deletions or insertions. Most changes are due to point mutations. Most of the alterations in the amino acid sequences which are due to those mutations are located in those regions which were found to be highly variable. Differences in the secondary structures (Figure 13,14) were identified in regions V1, V2, V3, and V4, corresponding with antigenic epitopes I, II, IV, and VI, and in epitopes V and X which are located in slightly variable regions (76 and 56% of conserved amino acids, respectively). A further β turn alteration occurs in region C5 in a short hydrophilic environment which may be located between two transmembrane-spanning regions. That means that all alterations concerning the secondary structure of the envelope protein, with the exception of that in region C5, are located in the predicted antigenic determinants; epitopes III, VII, VIII, IX, and XI are conserved. Interestingly, some of the β turn alterations identified in isolates WMJ1 and WMJ2 are reversed in isolate WMJ3 (I, II, IV, and VI), even if there are further variations in the amino acid sequence (I and IV).

After cleavage of the leader peptide from precursor protein gp160, gp120 represents the highly glycosylated exterior part of the envelope protein complex of HIV. This polypeptide part is likely connected only via the ionic interactions of 14 positively charged amino acids in the carboxy-terminal region of gp120 (region C4) adjacent to the cleavage site with the negativity charged phosphate groups in the bilayer membrane. There are 18 conserved

cysteine residues dispersed over the gp120 sequence (Figure 13) which might also be involved in the complex formation of gp120 and gp41. A report describing the frequent loss of gp120 from the particles during purification and immunoelectron microscopy argues, however, against a direct involvement of disulfite bonds in the complex formation of gp120 and gp41.

The transmembrane polypeptide gp41 contains three stretches of hydorphobic amino acids. Directly after the cleavage site there are about 60 hydorphobic residues (amino acids 511 to 571) which are interrupted by a stretch of about 10 amino acids that are predicted to have a high surface probability and, in comparison with the surrounding residues, a higher hydrophilicity and flexibility and some alterations in β turns. From the patterns of other viral (121) or membrane proteins, e.g., bacteriorhodopsin, acetylcholine receptor, and erythrocyte band III, such regions are known to represent sequential transmembrane-spanning regions. A domain of hydrophilic, flexible, and charged amino acids (amino acids to 572 to 594) linked to the hydrophobic region may represent a hydrophilic anchor sequence.

A transmembrane region, which was suggested from amino acids 670 to 695 with its hydrophilic anchor has already been mentioned. In transmembrane regions TM2 and TM3 three charged amino acids (Arginine at positions 557 and 707 and glutamic acid at position 560) are identifiable. Those charged amino acids were also reported in other transmembrane regions in multi-spanning proteins and do not lead to disturbance of the overall hydrophobic character of a transmembrane region. Between TM2 and TM3 hydrophobic regions from the N terminus of gp41 are 90 amino acid residues (580 to 670) which should be facing the outer side of the lipid bilayer. Within this region are four conserved glycosylation sites and a good antigenic probability (epitope X). Alternatively, TM1 and TM2 may represent an apolar hydrophobic stretch of amino acids located outside the viral membrane and may play a role similar to the paramyxovirus fusion protein by penetrating into the membrane of the fusion partner. The same could occur if the C-terminal part of gp41, including or not including TM3, would be directed to the outer membrane or be in equilibrium between those two orientations.

This possibility is supported by the observation that antiserum gains a peptide from amino acids 725 to 752 gives cell surface labeling and is neutralizing. The predicted antigenic determinants are not affected by the discussed models.

4. **To determine the molecular and biologic properties of HIV-2 isolates from individuals in West Africa where HIV-2 appears to be less virulent compared with HIV-1 and some other strains of HIV-2.**

Serological studies first suggested that a new type of HIV-1 related virus might be present in West African human populations (87). It was suggested that this virus, then termed "HTLV-4" might be closely related to a similar virus isolated from African green monkeys, then terms "STLV-3" (87,88). Because of similarities in structure and antigenicity among HIV-1, HIV-2 and STLV-3 (also referred to as SIV, simian immunodeficiency virus), some relation might be expected between their nucleic acid sequences and phylogenesis. Restriction enzyme analysis showed that six out of six STLV-3_{AGM} isolates and two out of three HTLV-4 isolates were identical in 32 out of 32 restriction sites mapped. The third HTLV-4 isolate was identical to the others in 31 out of 32 mapped restriction sites. Of three STLV-3_{MAC} isolates obtained independently at the New England Regional Primate Center, one (isolate 251) was identical to the STLV-3_{AGM}/HTLV-4 consensus pattern in 23 out of 23 sites, and the others were identical to each other and similar to the consensus pattern in 21 out of 23 sites. These findings raised the suspicion that HTLV-4 and STLV-3 reported to have been isolated from West African people and monkeys could represent laboratory containments with SIV_{MAC} isolate from macaque monkeys housed in the NERPC (39).

To define more precisely the relation between different HTLV-4, STLV-3, HIV-2 and HIV-1 viruses, we performed a combination of restriction endonuclease cleavage mapping, molecular cloning, and nucleotide sequencing to analyze and compare different isolates of HTLV-4 and STLV-3. The DNA from two HTLV-4 isolates (HTLV-4/PK82, HTLV-4/PK289) and two STLV-3 isolates (STLV-3_{AGM}/K1, STLV-3_{AGM}/Mm 6324) was extracted from virally-infected

cell-pellets without interim cell cultivation. Two other virus isolates (HTLV-4/PK190, STLV-3_{AGM}/K1) were cultured and used to prepare RNA for a viral-specific partial complementary DNA clone. A viral-specific insert from one of these clones (HTLV-4/PK190, clone pV2) was used to probe recombinant lambda phage libraries prepared from DNA infected with HTLV-4/PK82 and HTLV-4/PK289. Full-length proviral DNA clones including flanking cellular sequences, were obtained for both HTLV-4 viruses. The restriction enzyme cleavage patterns of STLV-3_{AGM}/K1, STLV-3_{AGM}/Mm6324, HTLV-4/PK82, HTLV-4/PK289 and of the full-length lambda phage clones derived from the two HTLV-4 isolates were compared. Eight restriction enzymes (*NdeI, PstI, XbaI, BamHI, Hind III, SstI, Bgl II, EcoRI*) were used. The restriction cleavage patterns of these isolates and proviral clones were identical in 19 out of 19 sites mapped, and corresponded to the HTLV-4/STLV-3_{AGM} consensus map.

The DNA from proviral or cDNA clones of HTLV-4/PK82, HTLV-4/PK190 and STLV-3_{AGM}/K1 was sequenced to define their relation to each other and to HIV-2 and other STLV-3 isolates. For HTLV-4/PK82, the regions sequenced included the entire long terminal repeat (LTR), *gag* p16 equivalent, central region, envelope, and 3' open reading frame (5,391 base paris). For HTLV-4/PK190 and STLV-3_{AGM}/K1, the sequence included R and U3 LTR regions and portions of 3' *orf* (850 and 1,160 base paris, respectively). A summary of this nucleotide sequence information is shown in Figure 17 and has been submitted to EMBL/GenBank under accession code number Y00283.

From this comparison of the sequences of HTLV-4/PK190 and three different STLV-3_{AGM} (K78, K6W, K1), none differ from HTLV-4/PK82 by more than 1.1%. In the envelope gene, which varies most widely in independent isolates of HIV-1, the three STLV-3_{AGM} sequences vary from HTLV-4/PK82 by 0.8% or less. But the virus STLV-3_{MAC}/Mm142-83 differs from both HTLV-4/PK82 and STLV-3_{AGM}/K6W by about 4.7% in envelope. Figure 18 depicts the distribution of deduced amino acid changes present in the envelopes of different STLV-3_{AGM}, STLV-3_{MAC}, and HIV-2 viruses with respect to HTLV-4/PK82. Compared to HTLV-4/PK82, there

were 10 amino acid differences in STLV-3_{AGM}/K6W, 14 in STLV-3_{AGM}/K78, and 65 in STLV-3_{MAC}/Mm142-83. All three STLV-3 and both HTLV-4 sequences (Figures 17 and 18) contain an identical stop codon in the envelope transmembrane coding region. Also, STLV-3_{MAC}/Mm142-83 compared to HTLV-4/PK82 contains clustered nucleotide and amino acid substitutions and a three-base-pair (bp) insertion. In the same regions of the viral envelope, HTLV-4 and STLV-3_{AGM} are identical. As a result of the nearly identical restriction patterns and nucleotide sequences of three different HTLV-4 and six different STLV-3_{AGM} isolates, and because of the failure to re-isolate such STLV-3-like viruses from humans, and in view the own isolation of a distinct HIV-2-related virus from a healthy Senegalese woman, we concluded that HTLV-4 and STLV-3_{AGM} are not independent laboratory isolates, but instead represent transmission of a single laboratory virus strain to multiple cultures. Furthermore, the various HTLV-4 and STLV-3_{AGM} isolates probably originated from cultures of STLV-3_{MAC} isolate 251, because the HTLV-14/STLV-3_{AGM} restriction site consensus pattern coincides precisely with isolate 251 in 32 out of 32 mapped sites, but is similar to STLV-3_{MAC} isolate 142 in only 27 out of 31 sites.

Despite showing that HTLV-4 was not a *bone fide* virus isolate, it was clear from clinical and serological data that West African human populations were infected with an HIV related virus that was more similar to SIV than to HIV-1. We thus attempted to isolate retroviruses from the peripheral blood of four healthy seropositive Senegalese prostitutes. Mononuclear cell preparations were cocultivated with uninfected normal donor lymphocytes. After 1-2 weeks, transient reverse transcriptase activity was detected in culture supernatants from three individuals, and virus from one of these cultures (isolate ST) was transmitted successfully to four immortalized T-cell lines, SUPT1, HUT78, H9, and CEM174. On the basis of its ultrastructural morphology, particulate reverse transcriptase activity, T4 cell tropism and antigenic reactivity by Western blot hybridization analysis, this virus was related to both HIV-2 and STLV-3 (SIV). Southern blots of virally-infected cellular DNA from these ST cultures using

DNA probes specific for HIV-2 and for HTLV-4/STLV-3, however, clearly showed the ST isolate to be more related to HIV-2 than to STLV-3/HTLV-4, as shown in Figure 19. The proviral DNA of the Senegalese isolate hybridized preferentially to HIV-2 and the restriction pattern of ST was unique. Moreover, two related yet distinguishable viral genotypes were found in the ST isolate, reflecting a genotypic heterogeneity STLV-3 (SIV) virus was not detected in any of the four Senegal cultures. These data demonstrate the existence of HIV-2 in Senegal and suggest that West African individuals seropositive for HTLV-4 are indeed infected with a retrovirus but that this belongs to the HIV-2, not STLV-3 (SIV), group.

Although human immunodeficiency virus type 1 (HIV-1) is the etiologic agent of epidemic AIDS in Central Africa, Europe, the United States, and most countries worldwide, many cases of AIDS in West Africa have been attributed to the related retrovirus, human immunodeficiency virus type 2 (HIV-2) (26-28). Although it is clear that HIV-2 in many instances causes fatal immunodeficiency (26-28), recent studies (41,86,107) have suggested that many West African individuals as a group, when infected with HIV-2, may have less severe immunodeficiency than individuals infected with HIV-1. These studies involved widespread serologic testing of West African populations by immunoblot and radioimmunoprecipitation (RIP) assays with viral-specific target antigens derived from cultures of HTLV-4/PK82 (87), a virus strain shown above and elsewhere (76,91,94) to represent a simian retrovirus termed SIV_{MAC} (simian immunodeficiency virus of macaques). In our studies described above, we characterized an HIV-2 related virus (HIV-2/ST) from a healthy prostitute in Senegal, West Africa which represented the first independent virus isolation from this region. The biologic and genetic properties of this virus distinguish it from other known immunodeficiency-associated retroviruses, and raise the possibility that such strains of HIV-2 may be responsible for altered disease presentations in naturally infected human populations.

HIV-2/ST was isolated from peripheral blood mononuclear cells (PBMC) of patient ST. Virus was transmitted to four immortalized T-cell lines (Hut78, H9, SupT1, and CEMx174) using

repeated PEG (polyethylene glycol) precipitations of primary culture supernatants in an attempt to isolate and amplify viruses with potentially attenuated virulence. The DNA genome of HIV-2/ST in these cultures hybridized at low and high stringency to the prototype HIV-2/ROD probe (Figure 20) whereas SIV_{MAC} DNA hybridized to the HIV-2/ROD probe only under conditions of low stringency. Conversely, HIV-2/ROD and HIV-2/ST DNA hybridized with equal intensity to SIV_{MAC} probe, but only at low stringency, and a full-length HTLV-1 probe gave no hybridization. Two different HIV-2/ST producing cell lines derived from the transmission of cell-free virus from primary ST lymphocyte co-cultures exhibited very similar proviral restriction enzyme cleavage patterns but were distinguishable by polymorphic Xba I and BamHI sites (Figure 20). A recombinant DNA lambda phage library prepared from one of these cell lines, Hut 78/B12, which represented a biologically cloned high producer line, yielded three full-length proviral DNA clones that were evaluated by detailed restriction analyses. Comparative restriction mapping confirmed the genetic similarity between HIV-2/ST and HIV-2/ROD (10 of 25 restriction sites in common) and the dissimilarity between HIV-2/ST and SIV_{MAC} (2 of 36 restriction sites in common) (Figure 20). Furthermore, one of these proviral DNA clones (lambda JSP4-27), when transfected into T-lymphocyte cell lines, produced virus that was replication-competent and infectious and whose biological properties were identical to the parental virus isolate. Thus, HIV-2/ST is genotypically more related to HIV-2/ROD than to other primate viruses, and within the HIV-2/ST isolate there exist polymorphic genotypes some of which are fully replication competent.

The structural, antigenic, and functional characteristics of HIV-2/ST proteins were evaluated by immunoblot and RIP analyses (Figure 21). The putative major *gag* structural protein and precursor (p26/pr55) of HIV-2/ST were comparable in size to p26/pr55 of HIV-2/ROD, p27/pr55 of SIV_{MAC}, and p24/pr53 of HIV-1. Polymerase and endonuclease proteins of HIV-2/ST were tentatively identified as p64 and p34 proteins with equivalent counterparts in HIV-2/ROD, SIV_{MAC}, and HIV-1. The *env* proteins of HIV-2/ST revealed a 140

kD putative extracellular protein similar to that in HIV-2/ROD. However, in different subcultures of HIV-2/ST that were derived by single cell cloning of the original SupT1/LK001 cell line, the putative transmembrane proteins (TMP) were either 42 kD or 30 KD in size. Only smaller-sized transmembrane proteins were seen in HIV-2/ROD (36 kD) and SIV_{MAC} (32 kD) cultures. The HIV-2/ST viruses with large and small transmembrane proteins could be transmitted cell free, were replication competent, and retained the same size of TMP throughout passage in the same cell targets (SupT1 and CEMx174). This suggests that the observed differences in TMP sizes were due to nucleotide sequence differences in the expressed mRNAs, not to post-transcriptional modifications such as altered glycosylation patterns. Consistent with this suggestion is the finding of translational stop codons in HIV-2/ROD and SIV_{MAC} viral envelope genes which result in the expression of truncated TMPs (28). Antigenically, HIV-2/ST and HIV-2/ROD were similar to each other and to SIV_{MAC} in their reactivity with HIV-2 antisera and were cross-reactive with HIV-1 only in the major *gag* structural proteins (p24-p27). This was confirmed by bi-directional immunoblotting in which both HIV-1 and HIV-2 specific human antisera were used to probe HIV-1, HIV-2/ROD, and HIV-2/ST proteins (Figure 21). By RIP, we found *env* precursor proteins of approximately 180 or 170 kD in the different HIV-2/ST subcultures (reflecting the different sizes of their respective transmembrane proteins) and mature extracellular envelope proteins of 140 kD (Figure 21). By RIP, we also found that, like HIV-1, the HIV-2/ST external *env* glycoprotein bound directly to an epitope on CD4 recognized by OKT4A (Figure 2C), but not by OKT4 (112). Other experiments with fluorescence activated cell sorting analysis showed that infection of CD4 bearing cells by HIV-2/ST virus down-modulated the expression of CD4(T4) but not CD8(T8), again analogous with HIV-1 infection (111).

Because of the evidence suggesting that certain West African human populations infected with HIV-2 related viruses may have less severe immunodeficiency than do individuals infected with HIV-1, or some isolates of HIV-2 including HIV-2/ROD (41,86,107), we evaluated

HIV-2/ST virus for cytopathic and cell killing properties *in vitro*. We first studied the induction of syncytia by HIV-2/ST infected cells co-cultured with a panel of different CD4$^+$ cells. As shown in Figure 22, both HIV-1/IIIb and HIV-2/ROD, but not HIV-2/ST, produced large syncytia when virally-infected cells were co-cultured with uninfected CD4$^+$ cells. Hela-T4, H9, and Hut 78 indicator cells gave identical results. Syncytia-induction was inhibited by the monoclonal antibody Leu3a and no syncytia were formed by any HIV-1 or HIV-2 infected cells when co-cultured with CD4$^-$ cells. Cytopathicity was also assessed by direct determination of cell killing. HIV-2/ROD and two different isolates of HIV-1 (IIIb and BC) caused marked cell killing of SupT1 cells, PHA-stimulated normal donor lymphocytes (Fig. 4, C and D), as well as Hut78 and H9 cells (Figure 23). Equivalent amounts of HIV-2/ST showed no detectable cell killing activity against the same cell targets. Even 10- and 100-fold more HIV-2/ST virus than three other HIV-1 isolates (WMJ, RH, BC) led to only transient depression in cell counts. These cell-killing experiments were repeated three times with identical results, each time with the use of newly concentrated virus stocks. Syncytia induction that accompanied virus infection in these experiments was also scored and was dramatic in size and number in cultures infected by HIV-2/ROD and the four HIV-1 isolates but absent in cultures infected with HIV-2/ST. The virus stocks used for these experiments were obtained from the original uncloned HIV-2/ST infected line SupT1/LK001 and were infectious and replication-competent on the basis of their ability to infect, replicate in, and reinfect by cell-free passage CD4$^+$ cells. Furthermore, the cell targets shown in Figure 23, and the virus infected cells used for the syncytia induction assays in Figure 22, were comparably infected with HIV-2/ST, HIV-2/ROD, and HIV-1 based on viral-specific immunofluorescence, *in situ* hybridization, and supernatant particulate reverse transcriptase assays.

In previous studies of the infectivity and fusion activity of SIV_{MAC}, it was observed that a cell line (CEMx174) representing a somatic cell hybrid between a T-cell (CEM) and a B-cell (B721.174), but neither parental cell line itself, was exquisitely sensitive to virus infection and

syncytia induction (82). We therefore used this hybrid cell line (CEMx174) in cytopathicity assays with HIV-2/ST. CEMx174 cells were highly susceptible to HIV-2/ST infection and were permissive for permanent, high-titer virus production. Syncytia induction and cell-killing assays showed a modest cytopathic effect but again much less than for prototype strains of HIV-1 and HIV-2. To evaluate the basis for the attenuation of HIV-2/ST, we used *in situ* hybridization to analyze at the single cell level the rate of virus entry and early gene expression after cell-free viral transmission (Figure 24). Compared to prototype strains of HIV-1 and HIV-2, HIV-2/ST showed markedly delayed onset of viral RNA production in CEMx174 as well as in SupT1 and Hut 78 following exposure to cell free virions. However, once infected, cells expressed equivalent amounts of HIV-2/ST RNA as compared with HIV-2/ROD and HIV-1/IIIb, and produced equivalent amounts of mature virus as assessed by RT activity and viral structural proteins. This, and the fact that HIV-2/ST is less fusogenic, suggests that the defect in HIV-2/ST is at the level of virus entry, not expression.

Previous studies (3, 4, 12) have shown that viruses of the HIV-2 group may cause severe and fatal immunodeficiency. Viruses cultured from these diseased patients were readily isolated by standard lymphocyte co-culture techniques and showed peaks of supernatant RT activity and cytopathic effects within the first 2 weeks of culture, similar to HIV-1. We had considerable difficulty in establishing virus isolates from healthy Senegalese subjects. Out of four patients originally studied, RT activity in lymphocyte co-culture supernatants was detected in three, but in only one case was a virus (HIV-2/ST) successfully amplified and transmitted to permanently producing immortalized T-cell lines. Taken together with the cytopathicity and infectivity data, these results suggest that certain strains of HIV-2 may be less virulent than others and in turn be associated with a less severe or altered expression of immune deficiency. The molecular basis for the attenuated virulence of HIV-2/ST *in vitro* is not known. However, the demonstration that gp120-T4(CD4) binding is followed by down-modulation of cell surface T4(CD4) expression but not cell fusion, and the apparent delay in cell entry and

spread after cell-free viral transmission, suggest that changes in the viral envelope important for membrane fusion or other post-binding events are involved. We have also isolated from a single subject on two occasions a strain of HIV-1 that, like HIV-2/ST, was poorly transmissible and required repeated PEG precipitations of culture supernatants for successful amplification. Mullins and co-workers (118) have recently described a replication defective strain of feline leukemia virus (FeLV) whose variant *env*/LTR region is actually responsible for *enhanced* virulence and they have suggested that biologically important strains of human immunodeficiency viruses could have similarly altered genetic and biologic properties. This report of an attenuated form of HIV-2 is an example of such a naturally-occurring HIV variant.

The significance of different-sized *env* transmembrane glycoproteins in HIV-2/ST is unknown. Biologically-cloned cultures of HIV-2/ST with either large (42 kD) or small (30kD) TMPs were both infectious, replication-competent, and non-cytopathic as were isolates of SIV_{MAC} and HIV-2/ROD having only small TMPs. These data suggest that a full-length (gp 42) TMP is not required for any of these biological functions, although the carboxyterminus of the TMP reading frame, which is highly conserved among HIV-1, HIV-2, and SIV_{MAC}, could potentially encode a separate protein with distinct biologic function.

Considerable controversy has surrounded the question as to what types of human immunodeficiency viruses, besides HIV-1, are present in West Africa. HIV-2/ST was shown by differential nucleic acid hybridization, restriction mapping, and analysis of core proteins to be considerably more related to the prototype HIV-2/ROD virus than to HIV-1 or SIV_{MAC}. There was no evidence in any of the primary or secondary cultures of HIV-2/ST, or in cultures from the other three Senegalese subjects similarly studied, for the presence of an SIV_{MAC}-like virus. Because of the near identity among isolates of HTLV-4 and SIV_{MAC}-251 at the genetic level, the failure to reisolate similarly conserved SIV_{MAC}-like viruses from West Africans, and the general finding of genotypic diversity among independent isolates of HIV-1, HIV-2, and SIV_{MAC}, we conclude that the HIV-2 viruses typified by HIV-2/ROD, HIV-2/SBL6669, and HIV-2/ST,

along with HIV-1, represent the naturally occurring human immunodeficiency viruses in West Africa.

As part of our studies involving the genetic and biologic characterization of HIV-1 and HIV-2 viruses, we also focused attention on unique features of central region genes that distinguish HIV-1 from HIV-2. Nucleotide sequence comparisons of HIV-1, HIV-2 and SIV_{MAC} demonstrated an overall highly conserved genomic organization, represented by LTR-gag-pol-vif-central region-env-3'orf-LTR (3,22,63,114,130,131,139,156). HIV-2 and SIV_{MAC} are approximately 70% homologous at the nucleotide sequence level and each is approximately 45% similar to HIV-1 (22,63). Interestingly, there is an additional open reading frame in the genomes of HIV-2 and SIV_{MAC} that is missing in HIV-1. This open reading frame, termed X, is situated in the central viral region between vif and vpr genes and partially overlaps the vif open reading frame on its 5' end. The deduced amino acid sequence of X predicts a protein of 112 amino acids and 14.5kD calculated molecular weight and is conserved in 94 of 112 amino acid residues between HIV-2 and SIV_{MAC} (Figure 25). In this study we chemically synthesized HIV-2 and SIV_{MAC} X specific oligopeptides, raised heterologous antisera to these peptides in rabbits, and expressed the SIV_{MAC} X open reading frame as TrpE fusion proteins in *E. coli.* With these reagents, we demonstrated that X encodes a novel retroviral protein uniquely present in HIV-2 and SIV_{MAC} and provided evidence for its expression and immunogenicity *in vivo.*

Two synthetic peptides encompassing 19 residues from the N-terminus of the deduced X amino acid sequences of HIV-2/ROD and SIV_{MAC}/PK82 (underlined in Figure 25), were synthesized and used to generate rabbit immune sera. The peptides were prepared by solid phase methodology on an Applied Biosystems 430-A peptide synthesizer using a paramethylbenzhydrolamine resin (90). Synthetic peptides were coupled to keyhole limpet hemocyanin (KLH) at the carboxyterminus, after the addition of two glycine spacers and a cysteine residue. New Zealand white rabbits were immunized with 1mg of KLH-coupled

peptides emulsified 1:1 in Freunds complete adjuvant, boosted twice at two week intervals with the same peptides mixed 1:1 with Freunds incomplete adjuvant, and bled six weeks after the first immunization to collect immune sera. A peptide ELISA with non-KLH-coupled oligopeptides as antigen was performed to confirm high titer antipeptide antibodies in these sera (47).

The SIV_{MAC} X orf was also expressed as a fusion protein in *E. coli*. An Xho-II/HindIII fragment containing the entire SIV_{MAC} X open reading frame minus the first six nucleotides (SIV_{MAC} clone PKE 102, ref. 26) was subcloned into the bacterial expression vector pATH (152), immediately adjacent and in-frame with the bacterial Trp-E gene (pATH-X, Figure 26A). *E. coli* (HB101) transformed with pATH-X were treated in mid-log phase growth with Indole acrylic acid to induce the trp operon (116). Cell extracts were subjected to polyacrylamide gel electrophoresis, and a predominant protein of approximately 50kD, consistent with the predicted molecular weight of a TrpE/X fusion protein, was identified by Coomassie blue staining of sodium dodecyl sulfate polyacrylamide gels. A second plasmid construct was prepared by deleting an internal NruI/EcoRI fragment in pATH-X, which removed 95% of the TrpE coding sequence (Figure 26A). This plasmid retained only 16 amino acid residues derived from the TrpE gene on the N-terminus and resulted in an expressed X fusion protein of 15kD (delta TrpE/X). Both the TrpE/X fusion protein and the TrpE deleted X fusion protein comprised approximately 20% of the total cellular protein as judged by Coomassie blue staining of gel electrophoresed bacterial lysates. Sera from rabbits immunized with the SIV_{MAC} and HIV-2 X oligopeptides (Figure 26B, panels B and D, respectively) reacted strongly with these lysates on Western blots (TrpE/X lanes 2; delta TrpE/X lanes 3), but no reactivity was detectable with the corresponding pre-immune sera (panels A and C), or with lysates of cultures transformed with the non-recombinant vector (lanes 1). An anti-TrpE antiserum, used for control, detected both the 37kD TrpE protein alone as well as the 50kD TrpE/X fusion protein, but did not recognize the 15kD TrpE deleted X protein (panel E).

In order to identify the putative X protein in SIV_{MAC} and HIV-2 infected cells and cell-free virions, immunoblotting techniques were used to probe viral preparations for reactivity with the rabbit immune sera. Figure 27A shows the Western blot patterns of viral preparations from cultures infected with HIV-1 (lanes 1), HIV-2 (lanes 2), SIV_{MAC} (lanes 3), as well as mock-infected cells (lanes 4). A polyclonal human anti-HIV-1 serum (RR) and a polyclonal human anti-HIV-2 serum (ST) were used for control to confirm the presence of virus specific antigens. Probing with the X-specific rabbit immune sera (anti-SIV_{MAC}, panel B; anti-HIV-2, panel D) resulted in the identification of a 14 and 12kD viral protein in HIV-2 and SIV_{MAC} viral preparations, respectively, but not in similarly prepared preparations of HIV-1 or virally-uninfected control cultures. Similar results were obtained using whole cell lysates infected with these viruses. No reactivity was detected with the preimmune sera (panels A and C). Interestingly, sera from rabbits immunized with the HIV-2 X peptide recognized only the homologous HIV-2 X protein (panel D), while the SIV_{MAC} X peptide antisera recognized both the HIV-2 and the SIV_{MAC} X protein (panel B). The SIV_{MAC} X protein appeared to be slightly smaller as compared to HIV-2 and was reproducibly present in considerably less amounts. The significance of these differences is presently under study.

In order to determine whether the X protein was actually virion-associated, HIV-2/ST virus was first concentrated by ultracentrifugation and then further purified on a 20% to 60% continuous sucrose gradient prior to Western blot analysis. As shown in Figure 27B, HIV-2 X peptide immune sera (D) detected the identical 14kD X-protein on immunoblots of banded HIV-2 virions. The intensity of this X specific band was equal to slightly greater as compared to cell lysates infected with comparable amounts of HIV-2. A human serum (ST) from a healthy HIV-2 infected individual, and the preimmune rabbit antiserum, failed to detect the X protein in the same antigen preparation. These data were confirmed with a second HIV-2 isolate (HIV-2/ROD).

Identification of the 12-14kD protein as the virally encoded X gene product was further

established by competitive adsorption assays (Figure 27C). SIV_{MAC} and HIV-2 X immune sera were adsorbed with nitrocellulose-immobilized lysates of *E. coli* transformed either with the pATH cloning vector or with the pATH-X recombinant plasmid. Adsorption of the SIV_{MAC} X immune sera with the 50kD TrpE/X fusion protein resulted in a complete loss of antibody reactivity with both HIV-2 (panel B',lane 2) and SIV_{MAC} (panel B', lane 3) viral preparations. Preadsorption of the same sera with *E. coli* lysates transformed with the pATH vector alone gave hybridization signals equal to that in Figure 27A, panel B (not shown). The reactivity of HIV-2 X peptide antisera (D') was only slightly diminished by preadsorption with the SIV_{MAC} TrpE/X fusion protein, as expected, since this antiserum had failed to recognize the heterologous X protein on previous immunoblots (compare Figure 27A, lane 3, panel D). Taken together, these results confirmed the 12-14kD proteins as the HIV-2 and SIV_{MAC} virally encoded X proteins.

To evaluate whether the SIV_{MAC} X protein synthesized in *E. coli* could be of diagnostic value, we examined a limited number of patient and normal control sera for reactivity with this protein (Figure 28). Four sera from healthy West African HIV-2 infected individuals (a-d), four sera from AIDS patients seropositive for HIV-1 (e-h), and two healthy control sera (i-j) were tested side-by-side for reactivity to the 50kD TrpE/X protein (lanes 2), the 15kD TrpE deleted X protein (lanes 3), and lysates of bacteria transformed with the non-recombinant pATH vector (lanes 1). Out of four HIV-2 antibody positive sera tested, two contained antibodies reactive with both the 50kD TrpE/X and the 15kD delta TrpE/X protein (panels a and c). These sera titered to a dilution of 1:50. Four HIV-1 positive sera and two HIV negative control sera did not react with these X specific fusion proteins (panels e-j). Bands corresponding to bacterial proteins other than the X fusion proteins were also observed, particularly with the West African sera. However, these were easily distinguishable from X-specific reactivities by their size and occurrence in the control antigen preparations (lanes 1).

In summary, the data demonstrate that HIV-2 and SIV_{MAC} viruses encode a novel

retroviral protein, termed X, which is expressed in virus cultures *in vitro* and in naturally-infected humans, *in vivo*. The molecular size of this protein, 12-14kD, approximates that calculated from the deduced amino acid sequence, although the actual structure and organization of the X gene and its protein, and the mode of X gene transcription in HIV-2 and SIV_{MAC} viruses, cannot be predicted. Similarly unknown at the present time is the function of the X protein in the life cycle of HIV-2 and SIV_{MAC}. The high degree of sequence conservation suggests a significant function for X in both viruses. The fact that HIV-1 lacks the X gene entirely, yet causes disease in infected individuals, implies that X gene expression is not required for induction of cytopathicity *in vitro* or *in vivo*. A search for X-related protein sequences in the NBRF protein data bank identified several proline-rich regions present in retroviral proteins, adenovirus, papillomavirus, human collagen, and a group of human phosphoproteins to share homology with the carboxyterminus of the X gene. In addition, a probable nuclear antigen of Epstein-Barr virus (9) showed 10 of 14 conserved amino acids with a region of X just 5' of the polyproline tract ($GRG^{H}_{R}G^{P}_{R}GG^{W}_{G}RPG^{P}_{A}P$). The significance of these homologies is under study.

The presence of antibodies specific for X in some HIV-2 infected individuals demonstrates both its immunogenicity and *in vivo* expression. It will be important to determine if X specific reagents will be generally useful in distinguishing between HIV-2 and HIV-1 infection, whether expression of X *in vivo* influences viral pathogenesis and clinical outcome, and whether titers of X specific antibodies will provide prognostically important clinical information. The availability of X specific reagents that we have developed as part of the current DoD contract should facilitate experiments addressing these questions.

5. **To determine the molecular basis for attenuation of cytopathicity in certain strains of HIV-2.**

HIV-1 and HIV-2 represent two distinct groups of human immunodeficiency viruses known to cause AIDS in infected individuals (12,26,27,57,125). While HIV-1 is the causative

agent of epidemic AIDS worldwide, HIV-2 has generally been restricted to West Africa. Numerous isolates of HIV-1 and HIV-2 have been obtained and their biological and molecular properties characterized (2,12,27,63,215,167). Nucleotide sequence analysis shows that HIV-2 is only distantly related to HIV-1 (63), while it is more closely related to two primate retroviruses, SIV_{MAC} and SIV_{SM}, which cause an AIDS-like disease in captive macaques (22,35,77). Although genetically divergent, prototype HIV-1 and HIV-2 viruses share similar overall genomic organization and have similar biological properties, which include their propensity for rapid genetic change (52,135,167), their host cell tropism, their cytopathic effect on T-cell cultures and peripheral blood mononuclear cells *in vitro*, and their ability to form syncytia with CD4-bearing target cells (27,125). In fact, the majority of HIV-1 and HIV-2 strains isolated from patients with immunodeficiency disease have been shown to cause cell fusion and the formation of multinucleated giant cells in culture. This represents a hallmark of productive viral infection in tissue culture and accounts for the profound *in vitro* cytopathic effects of HIV (73,103,165). In contrast to HIV-1 and HIV-2 isolated from clinically ill individuals, natural infection of African green monkeys by SIV occurs with high prevalence in East and West Africa but does not lead to immunodeficiency. In addition, there is growing evidence that the biologic properties of HIV-1 *in vivo* may differ significantly from tissue culture adapted strains. Thus, in order to characterize HIVs and SIVs as they exist *in vivo*, our goals were: (i) to characterize by genetic and biologic analysis an HIV-2 strain shown to have attenuated virulence *in vitro* and *in vivo*; (ii) to develop PCR approaches that would allow, for the first time, the direct identification, cloning, and sequence analysis of SIV and HIV from uncultured blood of West African green monkeys and humans (all previous clones of SIV_{AGM} have been derived from East African green monkeys); (iii) to obtain by direct lambda phage cloning full-length transfection competent clones of HIV-1 from uncultured human brain tissue.

In contrast to the prototype pathogenic HIV-1 and HIV-2 viruses, we and others have recently isolated less cytopathic strains of HIV-1 and HIV-2 which exhibit markedly different

biological properties. These particular isolates cause little or no cell death in susceptible target cells, fail to induce cell fusion with CD4-bearing immortalized T-cell lines, exhibit a restricted host cell tropism with a preference for PBMC and/or macrophages, and are often derived from asymptomatic individuals. While their *in vitro* biological differences were well-documented, the genetic changes responsible for their attenuated phenotype were not understood. In order to elucidate determinants of HIV pathogenicity, we have thus begun to molecularly dissect a non-fusogenic and non-cytopathic HIV-2 isolate, termed HIV-2/ST, which was obtained from a healthy Senegalese prostitute as part of this contract. Although this virus replicated to high titers in tissue culture, it infected cells at a slower rate compared to cytopathic strains of HIV-1 and HIV-2, and caused little or no cell killing and fusion. This was the case despite the fact that its external envelope glycoprotein was cleaved correctly, transported to the cell surface, and shown to bind to a specific epitope on CD4, which was recognized by OKT4a but not OKT4 antibodies. HIV-2/ST therefore appeared to bind to the CD4 molecule analogous to other HIVs, but it failed to fuse with CD4-bearing target cells suggesting that its infectivity was greatly retarded at the level of cell entry (see section 4 above).

Since HIV "isolates" generally represent complex mixtures of genotypically-distinct viruses, and since the biological phenotype of any HIV culture depends on the sum of the properties of each genotypic variant (61,135), we first attempted to isolate a molecular clone, which was both transfection-competent and representative of the *in vitro* properties of its parental virus. We therefore obtained three full-length proviral clones (λJSP4-27, λJSP4-32, and λJSP4-34) from a genomic library of a biologically-cloned high producer cell line, termed ST/B12, and subsequently transfected them into the neoplastic T-cell lines SupT1 (147) and CEMx174 (82,138). Reverse transcriptase activity was detected in supernatants of cultures transfected with λJSP4-27 as early as five days post-transfection, while λJSP4-32 and λJSP4-34 transfected cultures revealed no signs of viral replication, indicating that these proviruses were replication-defective. Immunofluorescence analysis further confirmed the

presence of virus expressing cells in λJSP4-27 transfected cultures, but failed to identify virus mediated cell fusion. Western blot analysis of purified JSP4-27 virions demonstrated a protein profile similar to that of the parental HIV-2/ST virus. To facilitate subsequent transfection experiments and to allow the direct comparison of JSP4-27 to other transfection-competent HIV-2 plasmid constructs, we sub-cloned the proviral insert of λJSP4-27 into the plasmid vector pSP65 (Figure 29A).

To test whether the transfection-derived JSP4-27 virions were infectious, filtered supernatants of plasmid-transfected cultures were transmitted to uninfected SupT1 and CEMx174 cells. The results showed that cell-free transmission of JSP4-27 virions was readily and reproducibly demonstrable. However, infection and spread in culture, particularly in SupT1 cells, occurred slowly and with considerable delay. These results were confirmed and extended in comparative studies with a transfection-derived, cytopathic HIV-2/ROD strain, termed SL1 (101). While transfection of the SL1 provirus resulted in >90% infected SupT1 or CEMx174 cultures within three to four days post-transfection, JSP4-27 transfected cultures reached only 10% infectivity in the same time period, which indicated the same reduced ability to spread in culture that had been observed for the parental HIV-2/ST virus (93). Similarly, the transfection-derived JSP4-27 cultures did not form syncytia upon cocultivation with several CD4-bearing T-cell lines, including SupT1, CEM, H9 and CEMx174 cells, whereas SL1 infected cultures produced numerous and large syncytia, as well as a profound cytopathic effect, with these same target cells (Figure 29B). These data thus showed that the JSP4-27 provirus was replication-competent and infectious, and exhibited the same non-fusogenic and non-cytopathic properties as previously described for the parental virus (93).

Having identified and characterized the biological features of a molecular clone of HIV-2/ST, we next sequenced its entire genome. The complete nucleotide sequence of the JSP4-27 provirus is depicted in Figure 30. The viral genome is 9,672 bp in length and exhibits an overall genomic organization of 5'LTR-gag-pol-central region-env-nef-3'LTR, which

is identical to that of other cytopathic HIV-2 and SIV_{MAC} proviruses. It contains all major open reading frames characteristic for HIV-2, including vpx which is present in HIV-2 and SIV_{MAC} viruses but not in HIV-1, and vpr which is present in HIV-1, HIV-2 and SIV_{MAC} but not in SIV_{AGM} viruses (55). Like other HIV/SIV proviruses, HIV-2/ST is flanked by LTR sequences which are known to regulate viral gene expression. Sequence comparison with other HIV-2 LTRs showed that regulatory elements, like the tata box, the polyadenylation site, core enhancer sequences, Sp1 binding sites, and the tat responsive region are all present in HIV-2/ST and that their sequences are highly conserved. The HIV-2/ST LTR is of similar length, and there are no major deletions or insertions which would distinguish it from the LTRs of other cytopathic HIV-2 viruses (data not shown).

Comparison of the deduced amino acid sequence of the HIV-2/ST reading frames suggested that they all encoded full-length and functional proteins, with the exception of the vpr gene. This open reading frame was found to contain an in-frame TAA stop codon which truncates the vpr protein prematurely after the first 32 amino acid residues. Since the JSP4-27 provirus is fully replication-competent, it can be concluded that the vpr gene product is not required for *in vitro* replication of HIV-2. This conclusion was confirmed by the biological analysis of a second vpr-deficient HIV-2 provirus independently constructed in our laboratory, as well as by the findings of others (37). Moreover, since vpr-deficient proviruses of HIV-2 are also cytopathic and fusogenic, it is unlikely that the lack of a functional vpr gene in HIV-2/ST is responsible for its attenuated phenotype.

Pairwise sequence alignments of JSP4-27 to other cytopathic strains of HIV-2 similarly revealed no genetic features unique for HIV-2/ST. Comparison of HIV-2/ST and HIV-2/ROD demonstrated an overall sequence divergence of 11%, which is within the expected range of genetic variability observed among geographically-distant isolates of HIV-2 (Figures 31,32). Three other recently reported HIV-2 viruses, HIV-2/ISY derived from a Gambian individual with AIDS (2,52), $HIV\text{-}2/NIH_z$ derived from an AIDS patient from Guinea Bissau (167), and

HIV-2/GH derived from an AIDS patient from Ghana (72,83) differ from HIV-2/ROD (Cape Verde Islands, refs. 27 and 63) by 11%, 12%, and 12%, respectively. Among all these viruses, the Senegalese HIV-2/ST virus was found to be most closely related to the Gambian isolate HIV-2/ISY, which shared 90% of its nucleotide sequence with HIV-2/ST.

Since infectivity, syncytia formation, and cell fusion are viral properties which are mediated by the viral *env* gene, we examined this gene in particular with respect to sequence differences unique for HIV-2/ST. An alignment of the deduced HIV-2/ST *env* sequence to those of six other cytopathic and fusogenic HIV-2 and SIV viruses is depicted in Figure 33. Overall, the size of the various *env* sequences compared is approximately the same. In contrast to other HIV-2 and SIV_{MAC} viruses, JSP4-27 contains no in-frame stop codon in its transmembrane envelope domain (Figure 33). This is consistent with the presence of a 43kD rather than a 32kD transmembrane glycoprotein on Western blots of JSP4-27 derived virions (Figure 29B), and is distinct from the protein profile of the HIV-2/ST parental strain which comprises a mixture of viruses with both full-length as well as truncated transmembrane proteins (93). Pairwise sequence alignment shows that the HIV-2/ST *env* sequence differs from other HIV-2 and SIV envelopes to the same degree as they differ from each other, with amino acid sequences varying between 16% and 30% (Figure 31). 25 of 32 cysteine residues, including 22 of 23 located in the extracellular domain, are conserved among all viruses, which indicate a highly conserved envelope structure. In addition, HIV-2/ST contains 28 potential N-linked glycosylation sites which are arranged in a pattern similar to that of other viruses, and which also include one highly conserved glycosylation site previously shown to be critical for HIV-1 infectivity (162). Finally, the HIV-2/ST *env* gene contains highly variable regions which correspond closely in distribution and size to similar hypervariable regions in the other *env* gene sequences.

Although a three-dimensional structure has not been determined for any HIV or SIV envelope glycoproteins, there are certain envelope domains whose functions have been

characterized by mutagenesis analysis. These include the putative CD4 binding domain (95,100), the envelope precursor cleavage site (110), and the viral fusion sequence (16,18,56,95). Since sequence changes in any one of these domains could alter the fusogenic properties of a virus (95), we analyzed the envelope sequence of HIV-2/ST for particular mutations in these areas. No changes, or only conservative amino acid changes were found in an envelope domain of JSP4-27 that corresponds to the HIV-1 envelope region previously identified to be involved in CD4 binding (100). HIV-2/ST also contained an apparently functional primary envelope precursor cleavage site, with a recognition sequence (RNKR) identical to that of three other fusogenic HIV-2 or SIV viruses (compare Figure 33). In contrast to these viruses, however, HIV-2/ST was found to differ in 2 of 16 highly conserved amino acid residues at the N-terminus of the transmembrane envelope glycoprotein which, as shown by site-directed mutagenesis, contains the viral fusion domain (16). The mutations include an alanine to threonine change involving amino acid residue 517 (position 12 after the cleavage site), and a serine to alanine change involving amino acid residue 521 (position 16 after the cleavage site). Only one other fusogenic HIV-2 virus, HIV-2/ISY, contained these same changes. However, this virus exhibited three additional mutations in this same envelope area (compare Figure 33). Since the fusion domain is generally highly conserved among cytopathic HIV and SIV viruses (16), we considered the possibility that the non-fusogenic properties of HIV-2/ST resulted from these mutations.

To determine if the observed amino acid substitutions in the HIV-2/ST envelope fusion region were likely responsible for the impaired cytopathic properties of this virus, we examined two fusogenic variants of HIV-2/ST, termed ST/24.1C and ST/24.2C (Figure 34). Both fusogenic strains were originally derived from a biologically-cloned sub-culture of HIV-2/ST, termed ST/24, which produced non-cytopathic and non-fusogenic virions biologically indistinguishable from the parental HIV-2/ST isolate. Following serial cell-free transmissions of ST/24 supernatant to uninfected SupT1 cells, large and numerous syncytia were observed

on two independent occasions, which indicated the emergence of fusogenic progeny virus in the culture. Two cell lines were subsequently established (ST/24.1C and ST/24.2C) and shown to produce virions with fusogenic and cytopathic properties similar to prototype HIV-1 and HIV-2 viruses. Moreover, these cell lines were confirmed to be infected with HIV-2/ST by Southern blot analysis, which revealed no changes in their BamHI, NheI, HindIII, and PstI cleavage patterns as compared to ST/24. In order to identify the molecular basis for the phenotypical change in these variants and to determine whether a direct mutation of the viral fusion sequence had occurred, we amplified the envelope fusion domain of these cultures using the polymerase chain reaction (136).

Two oligonucleotide primers (30 mers) were designed to allow the amplification of a 544 bp envelope fragment from virally-infected cellular DNA, which included the putative precursor cleavage site as well as the envelope fusion region (Figures 35 and 36). Both primers were synthesized according to the JSP4-27 sequence, however, sequence changes were introduced so as to accommodate a BamHI site in the 5' amplimer and a Pst I site in the 3' amplimer (primer 1: 5'AGAATTGGGGGATCCTAAATTGATAGAAGT 3'; and primer 2: 5'GCTATTTAATTTCTGCAGTTCATACATGTT 3'). Total genomic DNA of ST/24, ST/24.1C, ST/24.2C, as well as ST/B12 as a control, was amplified using these primers. 100 µl of reaction mixture contained 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatine, 200 mM of each deoxynucleotide triphosphate (dNTP), 10 pM of each primer, 2.5 U of Taq polymerase and 1 µg of high molecular weight DNA. Samples were subjected to 45 amplification cycles, consisting of a denaturing step at 94°C for 90 seconds, a primer-annealing step at 50°C for 90 seconds, and a primer extension step at 72°C for 135 seconds. Amplified envelope fragments were purified, cleaved with BamHI and PstI, and subsequently cloned into M13. Ten individual M13 clones per amplified DNA preparation were then isolated, and each clone was sequenced in the region, which corresponded to the N-terminus of the transmembrane envelope domain. An alignment of these sequences is depicted in Figure 35.

All 10 M13 clones derived from ST/B12 contained sequences identical to JSP4-27, which indicated that the PCR amplification procedure was reliable and did not cause frequent misincorporations of nucleotides in this particular DNA template. Sequence comparison of amplified fragments from cell line ST/24 demonstrated no differences among the individual M13 clones, but revealed 4 nucleotide point mutations between these ST/24 sequences and the JSP4-27 reference sequence. In fact, all ST/24 derived viruses, including the cytopathic and fusogenic ones, exhibited these same four nucleotide sequence differences, as well as the threonine and alanine substitutions previously identified in the envelope fusion region of JSP4-27. The results thus confirmed that the observed fusion sequence mutations were representative of all HIV-2/ST strains, regardless of their origin and biological phenotype. Interestingly, eight of nine M13 clones representing ST/24.1C and five of ten clones representing ST/24.2C contained additional point mutations, which predicted three amino acid sequence changes with respect to the ST/24 sequence (Figure 36). The presence of these mutations within the amplified material identifies the ST/24.1C and ST/24.2C viral strains as mixtures which comprise the parental ST/24 virus as well as additional genotypic variants. These newly-generated, genotypical variants must be responsible for the phenotypical changes seen in the ST/24.1C and ST/24.2C cultures. However, the genetic changes which are of biologically significance do not appear to occur in the envelope fusion domain.

While these studies were in progress, the biological properties of the HIV-2/ST envelope gene products were also analyzed in a eukaryotic expression system (115). Vaccinia virus expressed JSP4-27 envelope glycoproteins were compared to that of prototype HIV-1 and HIV-2 viruses. While the processing, expression, and transport to the cell surface appeared to be unaltered, vaccinia virus-produced JSP4-27 envelope glycoproteins failed to form syncytia with CD4-bearing Hela cells. Since no other HIV-2 proteins were produced in this system, these results strongly suggested that the JSP4-27 envelope glycoproteins were primarily responsible for its non-fusogenic and non-cytopathic phenotype (115). Based on these results

as well as the sequence data, we conclude that the non-fusogenic and attenuated properties of HIV-2/ST are indeed a function of its envelope gene products, although, the causative genetic defect appears not to involve mutations within the envelope fusion domain.

There are several mechanisms other than a direct alteration of the fusion sequence that could result in the biological changes observed in HIV-2/ST. These include mutations that affect envelope/CD4 interactions, mutations that cause differences in envelope glycosylation, mutations that require additional cell surface molecules to facilitate virus-cell fusion, as well as mutations that reduce the stability of envelope glycoprotein complexes on the cell surface. In fact, several naturally-occurring as well as genetically-engineered immunodeficiency viruses have been described, that are altered in their fusogenic or cytopathic properties because of one of these mechanisms. For example, possible differences in the binding affinity of HIV and SIV envelope glycoproteins to the CD4 receptor have been suggested by the finding that 25-fold more soluble CD4 is necessary to block infectivity of prototype HIV-2 viruses compared to HIV-1 (25). It is possible that the HIV-2/ST envelope glycoprotein binds the CD4 molecule with an even lower affinity, which would be expected to influence subsequent steps of viral entry, including membrane fusion and penetration. Another mechanism known to cause attenuation of virulence in naturally-occurring retroviruses involves differences in post-translational modifications of envelope glycoproteins. Mullins and coworkers showed that the pathogenic determinants of an immunodeficiency-causing FeLV-FAIDS virus were dependent on the processing of particular envelope oligosaccharides (126). Since HIV-2/ST differs in number and distribution of its potential N-linked envelope glycosylation sites from other cytopathic HIV-2's, and since size differences between the exterior envelope glycoproteins of fusogenic and non-fusogenic HIV-2/ST strains have been observed, a biologically-significant change in the sugar composition of the HIV-2/ST envelope cannot be excluded. Finally, a requirement of accessory molecules for virus-cell fusion represents still another potential mechanism to influence retroviral cytopathicity. Studies involving SIV_{MAC} recently revealed that

this virus has a restricted host-cell range which comprises only a subset of CD4+ T-cell lines (82,92). While highly infectious and cytopathic for HUT78 and H9 cells, SIV_{MAC} does not fuse with CD4-bearing SupT1 cells. Moreover, SIV_{MAC} infects SupT1 cells only with considerable delay. It is thus conceivable, that SIV_{MAC} requires a surface molecule(s) in addition to CD4 to establish a productive infection in certain human T-cell lines. Since its infection kinetics and lack of cytopathic effect in SupT1 cells very much resemble those of HIV-2/ST, it is not unreasonable to speculate that HIV-2/ST might similarly require an additional cell surface molecule(s) for efficient cell fusion or penetration.

The availability of cytopathic variants of HIV-2/ST will be instrumental for future experiments designed to define the exact molecular determinants involved in HIV-2/ST attenuation. Molecular clones representing the fusogenic and cytopathic HIV-2/ST strains are expected to exhibit much less genetic divergence with respect to JSP4-27 than do unrelated HIV-2 viruses, like HIV-2/ROD or HIV-2/ISY. Therefore, a comparative sequence analysis is more likely to identify biologically-important differences, and the construction of chimeras between attenuated and cytopathic clones will be greatly facilitated. The fact that cytopathic and fusogenic ST/24 mutants evolved by cell free passage on two independent occasions indicates the presence of strong selective pressures for cytopathic and fusogenic viruses *in vitro*. It is possible that similar pressures are also present *in vivo* which may favor the emergence of more virulent strains in certain HIV infected individuals over time.

6. **To develop PCR techniques for sensitive and specific amplification of novel HIV/SIV genomes using universal primers and "nested" amplification and to apply this to the detection and characterization of SIV in uncultured blood of West African green monkeys and people.**

Wild caught African green monkeys (AGM) from East and West Africa are infected with SIV viruses at a high prevalence rate (30-60%) in contrast to Asian macaques which are not naturally infected (35,85-88). Because of this, it has been speculated that AGMs could harbor

viruses that served as the progenitor of human immunodeficiency viruses acquired through cross-species transmission. Recent sequence analyses of a sooty mangabey virus has shown it to be most closely related to HIV-2 (77). To date, isolation and/or molecular characterization of SIV_{AGM} from West Africa has not been reported. Of additional interest is the fact that wild caught AGMs, unlike experimentally infected rhesus macaques, do not appear to develop disease as a result of SIV infection. Because a molecular analysis of SIV_{AGM} from West Africa could provide important information relevant to the evolutionary relationships of all HIV and SIV viruses, and because the biologic properties of these viruses are likely to be enlightening, we developed a PCR-based approach for cloning these viruses directly from uncultured AGM peripheral blood lymphocytes. Figures 37 and 38 illustrate the "nested" PCR technique that we developed which accomplishes this. Using two sequential 30 cycle rounds of amplification, first with an outer primer set and then by an inner set containing cloning sites for the M13 polylinker, we successfully cloned SIV_{AGM} from four out of four seropositive AGMs and from zero of two seronegative animals. Figures 39, 40 and 41 show the nucleotide sequences of the amplified regions of the polymerase gene and their relationship to SIVs from other simian species and subspecies. These results are important for the following reasons: (i) they demonstrate for the first time the utility of the generic (universal) "nested" PCR approach for detecting and cloning novel groups of viruses whose sequence divergence from known viruses precludes conventional hybridization and cloning approaches; (ii) they identify the SIV strain infecting feral West African AGMs (sabeus) and show that this virus group is not closely related to HIV-1 or HIV-2; (iii) they provide a rapid method for obtaining molecular probes of novel viruses for conventional lambda phage cloning and other kinds of nucleic acid analyses.

7. **To molecularly clone full-length replication competent HIV-1 proviruses from uncultured human brain specimens for analysis of virus biology, molecular determinants of tropism and virulence, genomic organization, and gene structure-function relationships.**

All previous full-length replication-competent proviral clones of HIV-1 have been obtained from cell cultures of amplified virus, due to the low abundance of viral DNA in chronically infected individuals (145). Because such virus strains may be altered as a result of *in vitro* propagation (162), we sought to molecularly clone and genetically and biologically characterize full-length replication competent HIV-1 proviruses directly from uncultured human brain tissue of a patient with AIDS Dementia Complex (ADC). Objectives were (i) to generate replication competent proviral clones from uncultured human tissue thereby allowing an analysis of genome organization and gene structure-function relationships of virus not subjected to *in vitro* selection pressures, (ii) to generate transfection-derived genetically defined HIV-1 strains without interim cell culture for analysis of virus biological properties, replicative DNA intermediate forms, integration status, and the possible existence of defective and/or helper forms, (iii) to compare and characterize genotypic variation of proviral clones obtained by direct lambda phage cloning as compared to clones obtained by PCR amplification. For these studies, high molecular weight DNA from a brain specimen obtained at necropsy was subjected to lambda phage cloning and 10 HIV-1 proviral clones out of 8×10^6 recombinants were obtained (Figure 42). These proviral clones contained integrated and unintegrated forms, forms with one or two LTRs, and genomes with large deletions and self-integrated, reversed LTR sequences. Eight HIV-1 proviral clones (in lambda) and 11 PCR derived clones from the same brain were sequenced in a 525 bp hypervariable envelope region. All 19 clones were highly related yet distinct with nucleotide variation between 0.1 and 1.1%.

These nucleotide sequence data derived from lambda and PCR derived HIV-1 molecular clones from two different frozen specimens of brain tissue that were processed one year apart prove conclusively that the full-length HIV-1 proviral clones described herein are authentic, and in fact, represent the first such clones from uncultured human tissue. Figures 43 and 44 summarize the comparative nucleotide sequence data in the envelope region of the eight lambda clones and eleven PCR derived clones. 5/8 lambda clones and 4/11 PCR clones were

identical over this 525 base pair length. The other 3 lambda clones differed by 0.19 to 0.38% and the other 7 PCR clones differed from the predominant (consensus) form by 0.19 to 1.14%. Interestingly, two deletions and two stop codons were identified in the PCR clones but not in the lambda clones. Figure 44 shows the specific nucleotide and amino acid changes that were present in non-consensus clones, most of which represented nucleotide transitions. The biological significance of these changes are under study.

Four proviral clones in lambda were determined to be full-length by restriction mapping and two of these were shown to be transfection-competent in Cos-1 cells and replication competent in human T cells and monocytes after cell free passage. We have found in preliminary studies that these molecularly derived strains of HIV-1 from uncultured human brain replicate to higher levels in monocyte-macrophages than do other putative monocyte tropic viruses and therefore they represent important reagents for characterizing monocyte tropic viral determinants at a molecular level. The genomic organization, structure-function characteristics of specific gene products, and the biological properties of these HIV-1 proviral clones are under study and will be pursued aggressively in the future.

REFERENCES

1. Air, G.M., Proc. Natl. Acad. Sci. U.S.A., 78, 7639 (1981).
2. Albert, J., Bredberg, J., et al., AIDS Res. Human Retro, 3, 3-10 (1987).
3. Alizon, M., Wain-Hobson, S., Montagnier, L., Sonigo, P., Cell, 46, 63-74 (1986).
4. Alizon, M., et al., Nature, 312, 757-760 (1984).
5. Allen, J.S., et al., Science, 228, 1091-1093 (1985).
6. Asjo, B., et al., Lancet ii, 660-662 (1986).
7. Atassi, M.Z., Immunochemistry, 15, 909-936 (1978).
8. Baltimore, D., and Wolff, S.M., Eds., Confronting AIDS, (National Academy Press, Washington, D.C., 1986).
9. Baer, R., et al., Nature, 310, 207-211 (1984).
10. Barin, F., et al., Science, 228, 1094-1096 (1985a).
11. Barin, F., et al., Lancet ii, 1387-1389 (1985b).
12. Barre-Sinoussi, F., et al., Science, 220, 868-871 (1983).
13. Benn, S., et al., Science, 230, 949 (1985).
14. Berkelman, R.L., et al., Amer. J. of Med., 86, 761-770 (1989).
15. Bishop, J.M., and Varmus, H., In RNA Tumor Viruses, Ed. by Weiss, R. et al., pp. 301-327 (1985).
16. Bosch, M.L., et al., Science, 244, 694-697 (1989).
17. Brahic, M., et al., Nature, 292, 240-242 (1981).
18. Brasseur, R., et al., AIDS Res. and Human Retro., 4, 83-90 (1988).
19. Bruck, C., et al., Virology, 136, 20-31 (1984).
20. Buonagurio, D.A., et al., Science, 232, 980 (1986).
21. Centers for Disease Control, MMWR, 37, 33-35 (1988).
22. Chakrabarti, L., et al., Nature, 328, 543-547 (1987).
23. Cheng-Mayer, C., et al., Science, 240, 80-82 (1988).
24. Chou, P.Y. and Fasman, G.D., Biochemistry, 13, 222-245 (1974).
25. Clapham, P.R., et al., Nature, 337, 368-370 (1989).
26. Clavel, F., et al., New England J. Med., 316, 1180-1185 (1987).
27. Clavel, F., et al., Science, 233, 343 (1986).
28. Clavel, F., AIDS, 1, 135 (1987).
29. Clements, J.E., et al., Proc. Natl. Acad. Sci. U.S.A., 77, 4454-4458 (1980).
30. Coffin, J.M., Cell, 46, 1-4 (1986).
31. Coffin, J.M., J. Gen. Virol., 42, 1 (1979).
32. Cohen, G.H., et al., J. Virol., 49, 102-108 (1984).
33. Cortes, E., et al., New England J. Med., 320, 953-958 (1989).
34. Curran, J.W., et al., Science, 229, 1352 (1985).
35. Daniel, M.D., et al., Science, 228, 1201-1204 (1985).
36. Darlix, J-L., and Spahr, P-F., Nucl. Acid Res., 11, 5953-5967 (1983).
37. Dedera, D., et al., J. Virol., 63, 3205-3208 (1989).
38. Desai, S.M., et al., Proc. Natl. Acad. Sci. USA, 83, 8380-8384 (1986).
39. Desrosiers, R.C., et al., Nature, 327, 107 (1987).
40. Desrosiers, R.C., et al., J. Virol. 54, 552-560 (1985).
41. deThe, G., Proceedings of the Deuxieme Colloque Des Cent Gardes, Paris, (1987).
42. DiMarzo-Veronese, F., et al., Science, 229, 1402-1405 (1985).
43. Domingo, E., et al., Gene, 40, 1-8 (1985).
44. Dorner, A.J., Stoye, J.P., and Coffin, J.M., J. Virol., 53, 32-39 (1985).
45. Efstratiadis, A., et al., Cell, 21, 653-668 (1980).
46. Eisenberg, R.J., et al., J. Virol., 53, 634-644 (1985).
47. Emini, E. A., Jameson, B.A., Wimmer, E., Nature, 304, 699-703 (1983).

48. Evans, L.A., et al., Science, 240, 1522-1525 (1988).
49. Farabaugh, P.J., et al., J. Mol. Biol., 126, 847-863 (1978).
50. Fauci, A., Proc. Natl. Acad. Sci. U.S.A., 83, 9278 (1986).
51. Fisher, A.G., et al., Nature, 316, 262-265 (1985).
52. Franchini, G., et al., Proc. Natl. Acad. Sci. U.S.A., 86, 2433-2437 (1989).
53. Franchini, G., et al., AIDS Res. Hum. Retro., 4, 243-250 (1988).
54. Franchini, G., et al., Nature, 328, 543-547 (1987).
55. Fukasawa, M., et al., Nature, 333, 457-461 (1988).
56. Gallagher, W.R. Cell, 50, 327-328 (1987).
57. Gallo, R.C., et al., Science, 224, 500-503 (1984).
58. Gartner, S., et al., Science, 233, 215-219 (1986).
59. GenBank, Los Alamos National Laboratory, Los Alamos, New Mexico, 87545.
60. Gojobori, T. and Yokoyama, S., Proc. Natl. Acad. Sci. U.S.A., 82, 4198 (1985).
61. Goodenow, M., et al., J. AIDS, 2, 344-352 (1989).
62. Gunn, P.R., et al., J. Virol., 54, 791-797 (1985).
63. Guyader, M., et al., Nature, 326, 662-669 (1987).
64. Haase, A., et al., Science, 195, 175-177 (1977).
65. Hahn, B.H., et al., Nature, 312, 166-169 (1984).
66. Hahn, B.H., et al., Proc. Natl. Acad. Sci. U.S.A., 82, 4813-4817 (1985).
67. Hahn, B. H., et al., Nature, 330, 184-186 (1987).
68. Hahn, B.H., et al., Human Retroviruses, Cancer, and AIDS, p. 57-74 (1988).
69. Hahn, B., et al., Science, 232, 1548-1553 (1986).
70. Hahn, B.H., unpublished data.
71. Harper, M.E., et al., Proc. Natl. Acad. Sci. U.S.A., 82, 772 (1986).
72. Hasegawa, A., et al., AIDS and Hum. Retro., (1989).
73. Haseltine, J.G. and Sodroski, J., Acquired Immunodeficiency Syndrome, Eds., Gluckman, J.C., and E. Vilmer. (Elsevier) pp. 47-56.
74. Henderson, L.E., et al., Science, 241, 199-201 (1988).
75. Hirsch, V., et al., Proc. Natl. Acad. Sci. U.S.A., 83, 9754-9758 (1986).
76. Hirsch, V., Riedel, N., & Mullins, J.I., Cell, 49, 307-319 (1987).
77. Hirsch, V.M., et al., Nature, 339, 389-392 (1989).
78. Ho, D.D., et al., N. Engl. J. Med., 313, 1493-1497 (1985).
79. Holland, J., et al., Science, 215, 1577 (1982).
80. Hopp, T.P., and Woods, K.R., Proc. Natl. Acad. Sci. U.S.A., 78, 3824-3828 (1981).
81. Horsburgh, C.R., Jr. and Holmberg, S.D., Transfusion, 28, 192-195 (1988).
82. Hoxie, J.A., et al., J. Virol., 62, 2557-2568 (1988).
83. Ishikawa, K., et al., AIDS, 2, 383-388 (1988).
84. Kanki, P.J., et al., Lancet i, 1330-1332 (1985).
85. Kanki, P.J., Alroy, , & Essex, M., Science, 230, 951-954 (1985).
86. Kanki, P.J., et al., Science, 236, 827-831 (1987).
87. Kanki, P.J., et al., Science, 232, 238-242 (1986).
88. Kanki, P.J., et al., Science, 228, 1199-1201 (1985).
89. Kappes, J.C., et al., J. Virol., 62, 3501-3505 (1988).
90. Kent, S., and I. Clark-Lewis, Synthetic Peptides in Biology and Medicine, p. 29-58 (1985).
91. Kestler, H.W., et al., Nature, 331, 619-621 (1988).
92. Koenig, S., et al., Proc. Natl. Acad. Sci. U.S.A., 86, 2443-2447 (1989).
93. Kong, L.I., et al., Science, 240, 1525-1529 (1988).
94. Kornfeld, H., et al., Nature, 326, 610-613 (1987).
95. Kowalski, M., et al., Science, 237, 1351-1355 (1987).
96. Koyanagi, Y., et al., Science, 236, 819-822 (1987).
97. Krystal, M., et al., J. Virol., 45, 547 (1983).

98. Kuhnel, H., et al., Proc. Natl. Acad. Sci. U.S.A., 86, 2383-2387 (1989).
99. Laemmli, U.K., Nature, 227, 680-685 (1970).
100. Lasky, L.A., et al., Cell, 50, 975-985 (1987).
101. Lee, S.W., et al., Proceedings of the Colloque des Cent Gardes, Paris, France, pp. 11-17. (1988).
102. Li, W.H., Gojobori T., Nei, M., Nature (London), 292, 237 (1981).
103. Lifson, J.D., et al., Science, 232, 1123-1127 (1986).
104. Luciw, P.A., et al., Nature, 312, 760-763 (1984).
105. Lutley, R., et al., J. Gen. Virol., 64, 1433-1440 (1983).
106. Maniatis, T., et al., Molecular Cloning (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 326-328 (1982).
107. Marlink, R.G., et al., AIDS Res. Hum. Retroviruses, 4, 137 (1988).
108. Martinez, C., et al., Virology, 130, 539 (1983).
109. Maxam, A.M. and Gilbert, W., Methods Enzymol, 65, 499-560 (1980).
110. McCune, J.M., et al., Cell, 53, 55-67 (1988).
111. McDougal, J.S., et al., J. Immunol., 135, 3151 (1985).
112. McDougal, J.S., et al., Science, 231, 382 (1986).
113. Montelaro, R.C., et al., J. Biol. Chem., 259, 10539 (1984)
114. Muesing, M.A. et al., Nature, 313, 450-458 (1985).
115. Mulligan, M.J., et al., (submitted).
116. Nichols, B.P., and L. Yanofsky, Methods Enzymol, 101, 155-164 (1983).
117. O'Rear, J.J., and Temin, H.M., Proc, Natl. Acad. Sci. U.S.A., 79, 1230-1234 (1982).
118. Overbaugh, J., et al., Science, 239, 906-910 (1988).
119. Payne, S., et al., Virology, 161, 321-331 (1987).
120. Payne, S., et al., J. Virol., 61, 1266-1270 (1987).
121. Pellett, P.E., et al., J. Virol., 53, 243-253 (1985).
122. Perler, F., et al., Cell, 20, 555 (1980).
123. Petursson, G., et al., Lab. Invest., 35, 402-412 (1976).
124. Piot, P., et al., Science, 239, 573 (1988).
125. Popovic, M., et al., Science, 224, 497-500 (1984).
126. Poss, M.L., Mullins, J.I., and Hoover, E.A., J. Virol., 63, 189-195 (1989).
127. Queen, C.L. and Korn, L.J., Meth. Enyzmol., 65, 595-609 (1980).
128. Quinn, T.C., et al., Science, 234, 955 (1986).
129. Rabson A., Martin, M.A., Cell, 40, 477 (1985).
130. Ratner, L., et al., Nature, 313, 277-284 (1985).
131. Ratner, L., et al., Nature, 313, 636-637 (1985b).
132. Redfield, R.R., Wright, D.C., Tramont E.C., N. Engl. J. Med., 314, 131 (1986).
133. Robert-Guroff, M., et al., Nature, 316, 72-74 (1985).
134. Rosen, C.A., et al., Nature, 319, 555-559 (1986).
135. Saag, M.S., et al., Nature, 334, 440-444 (1988).
136. Saiki, R.D., et al., Science, 239, 487-491 (1988).
137. Salinovich, O., et al., J. Virol., 57, 71-80 (1986).
138. Salter, R.D., Howell D.M., and Cresswell, P., Immunogenetics, 21, 235-246 (1985).
139. Sanchez-Pescador, R., et al., Science, 227, 484 (1985).
140. Sanger, F., Nicklen, S., and Coulson A.R., Proc. Natl. Acad Sci. U.S.A., 74, 5463-5467 (1977).
141. Sarngadharan, M., et al., Science, 223, 506-508 (1984).
142. Schupbach, J., et al., Science, 224, 503-506 (1984).
143. Shaw, G.M., et al., Science, 227, 177 (1985).
144. Shaw, G.M., et al., paper to presented at the Second International Conf. on AIDS. Paris, France. 23 to 25 June 1986.
145. Shaw, G.M., et al., Science, 226, 1165-1171 (1984).

146. Shtivelman, E., Zakut, R., and Canaani, E., Proc. Natl. Acad. Sci. U.S.A., 81, 294-298 (1983).
147. Smith, S.D., et al., Cancer Research, 44, 5657-5660 (1984).
148. Srinivasan, A., et al., Blood, 69, 1766-1770 (1987).
149. Starcich, B., et al., Cell, 45, 637-648 (1986).
150. Steinhauer, D.A., and Holland, J.J., Ann. Rev. Microbiol., 41, 409-433 (1987).
151. Stowring, L., et al., Virology, 141, 311-318 (1985).
152. Tanese, N., M. Roth, and S. P. Goff, Proc. Natl. Acad. Sci., U.S.A., 82, 4944-4948 (1985).
153. Tersmette, M., et al., J. Virol., 63, 2118-2125 (1989).
154. Thormar, H., Barshatzky, M.R., Arnesen, K., Kozlowski, P.B., J. Gen. Virol., 64, 1427-1432 (1983).
155. Vilmer, E., (Elsevier) pp. 47-56 (1987).
156. Wain-Hobson, S., et al., Cell, 40, 9-17 (1985).
157. Weddell, G.N., et al., Proc. Natl, Acad. Sci. U.S.A., 82, 2618-2622 (1985).
158. Weiss, R., In RNA Tumor Viruses, Weiss, R., Teich, N., Varmus, H., Coffin, J., Eds. (Cold Spring Harbor Laboratory, Cold Spring, NY), pp.236-240 (1982).
159. Weiss, R.A, et al., Nature, 316, 69-72 (1985).
160. Westhoff, E., et al., Nature, 311, 123-126 (1984).
161. Wilbur, W.J., and Lipman, D.J., Proc. Natl. Acad. Sci. U.S.A., 80, 726 (1983).
162. Willey, R.L., et al., J. Virology, 62, 139-147 (1988).
163. Willey, R.L., et al., Proc. Natl. Acad. Sci. USA, 83, 5038-5042 (1986).
164. Wong-Staal, F., et al., Science, 229, 759-762 (1985).
165. Yoffe, B., et al., Proc. Natl. Acad. Sci. U.S.A., 84, 1429-1433 (1987).
166. Yu, X.F., et al., Nature, 335, 262-265 (1988).
167. Zagury, J.F., et al., Proc. Natl. Acad. Sci. U.S.A., 85, 5941-5945 (1988).
168. Zarling, D.A., and Temin, H.M., J. Virol., 17, 74-84 (1975).

APPENDIX A: FIGURES

Clones	LTR	Leader Sequence and Primer Binding Site	gag	pol	sor	NCR (tat)	envelope: Signal Peptide	envelope: Extra-cellular portion	envelope: Trans-membrane portion	3' orf	Total
Nucleotide Differences Compared to BH-10 (number of nucleotide changes/number of nucleotides sequenced)											
HAT-3	45/634	13/152	77/1285	46/1246	17/464	48/584	26/111	202/1443	87/1035	69/648	645/7489
	7%	8.5%	5.9%	3.7%	3.6%	8.2%	23.4%	14.0%	8.4%	10.6%	8.6%
ARV-2	30/634	14/152	86/1536	134/3045	31/609	49/584	26/111	164/1443	66/1035	52/648	582/9213
	4.7%	9.2%	5.6%	4.4%	5.1%	8.4%	23.4%	11.4%	6.3%	8.0%	6.3%
LAV-1a	10/634	7/152	46/1536	59/3045	2/609	11/584	2/111	32/1443	9/1035	13/648	144/9213
	1.6%	4.6%	3.0%	1.9%	0.3%	1.9%	1.8%	2.2%	0.9%	2.0%	1.5%
WMJ-1	nd	nd	38/1162[a]	nd	nd	nd	27/111	177/1443	62/1035	nd	nd
			3.3%				24.3%	12.3%	6.0%		
Amino Acid Differences Compared to BH-10 (number of amino acid changes/number of amino acids sequenced)											
HAT-3	–	–	28/428	22/415	10/154	–	11/37	103/481	45/345	37/216	256/2039
	–	–	6.5%	5.3%	6.4%	–	29.7%	21.4%	13.0%	17%	12.5%
ARV-2	–	–	32/512	51/1015	20/203	–	13/37	82/481	42/345	29/216	239/2593
	–	–	6.3%	5.0%	9.8%	–	35.1%	17.0%	12.2%	12.2%	9.2%
LAV-1a	–	–	16/512	21/1015	0/203	–	2/37	14/481	5/345	8/216	58/2593
	–	–	3.1%	2.1%	0%	–	5.4%	2.9%	1.4%	3.7%	2.2%
WMJ-1	–	–	12/387[a]	nd	nd	–	13/37	90/481	29/345	nd	nd
	–	–	3.1%			–	35.1%	18.7%	8.4%		

Figure 1. Sequence Comparison of Five Independent HIV-1 Proviral Clones

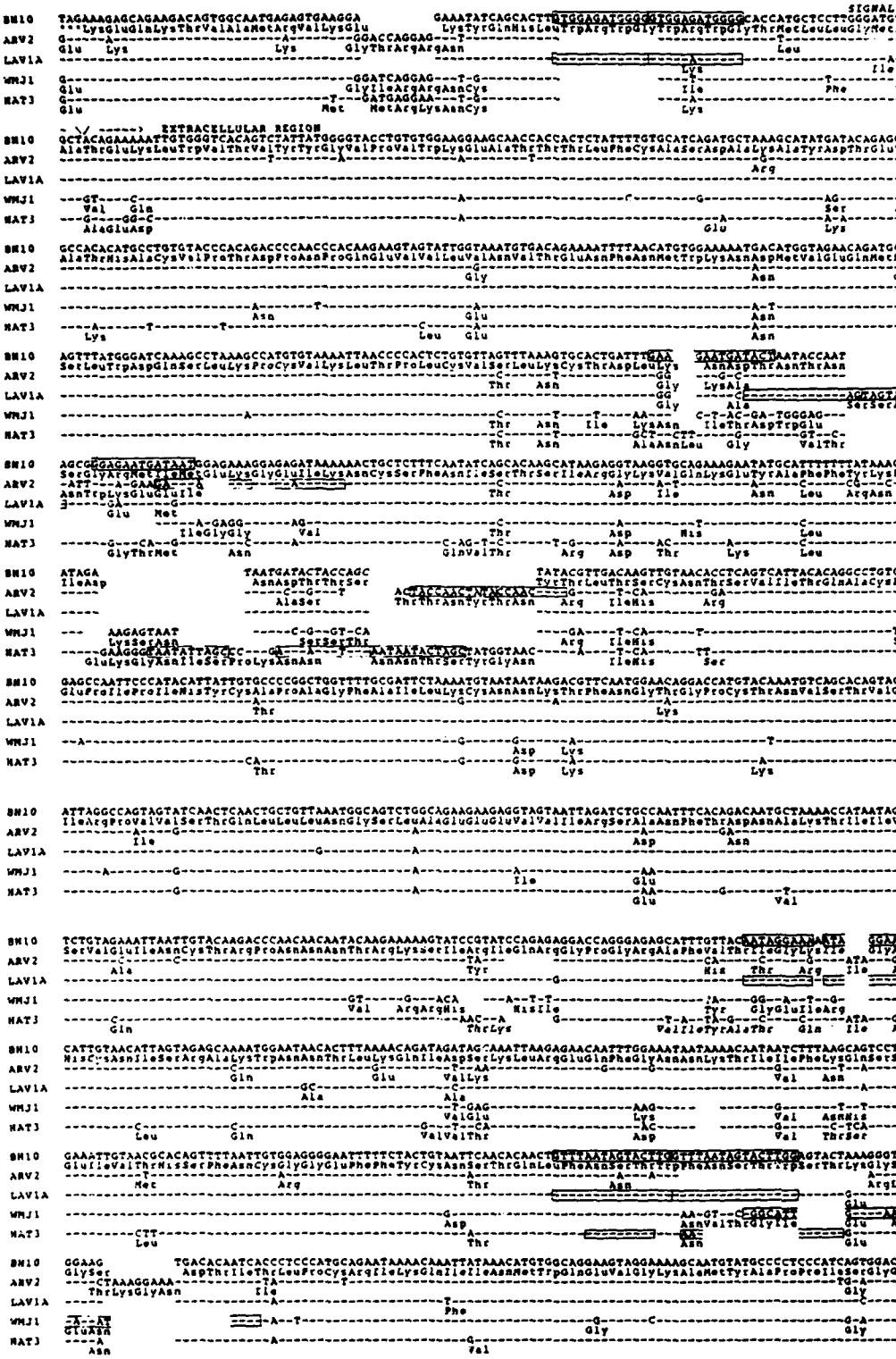
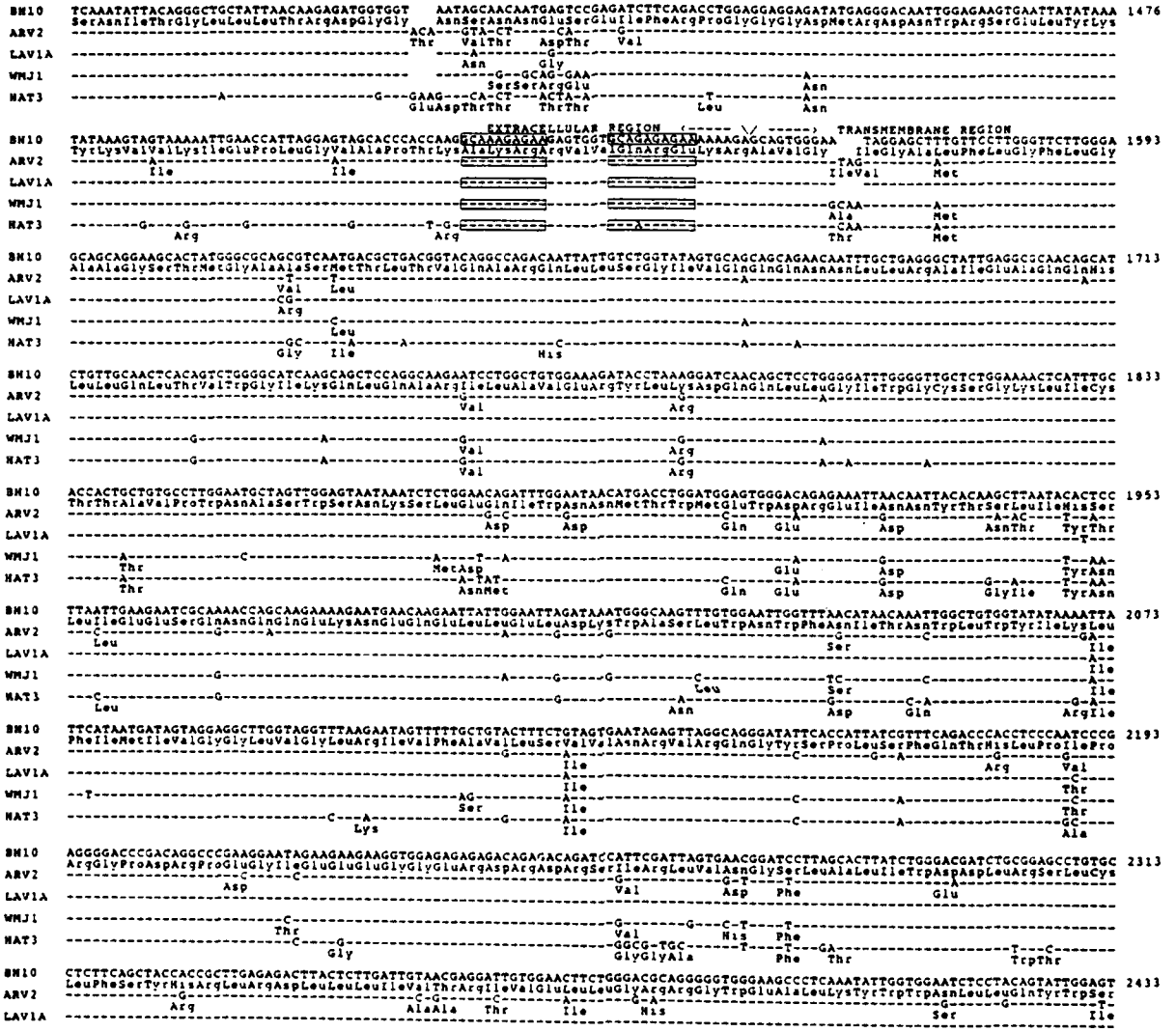


Figure 2. Nucleotide Sequences of the Entire Envelope Genes of Five Independent HIV-1 Proviral Clones. Sequence information for BH-10 (HTLV-IIIb), ARV-2, and LAV-1a was obtained from GENBANK. Sequences for WMJ-1 and HAT-3 were determined as part of the present study. Alignment of the sequences was performed pairwise with the assistance of NUCALN. Numbering of nucleotides is from the first nucleotide of BH-10 following the TAG stop codon, and genomic regions corresponding to the signal peptide, extracellular glycoprotein (gp 120), and transmembrane glycoprotein (gp 41) are shown. *Dashes indicate nucleotide identify with BH-10 and spaces* indicate the absence of that nucleotide. The boxed nucleotides comprise adjacent perfect or imperfect direct repeats.

Figure 2 (cont.)



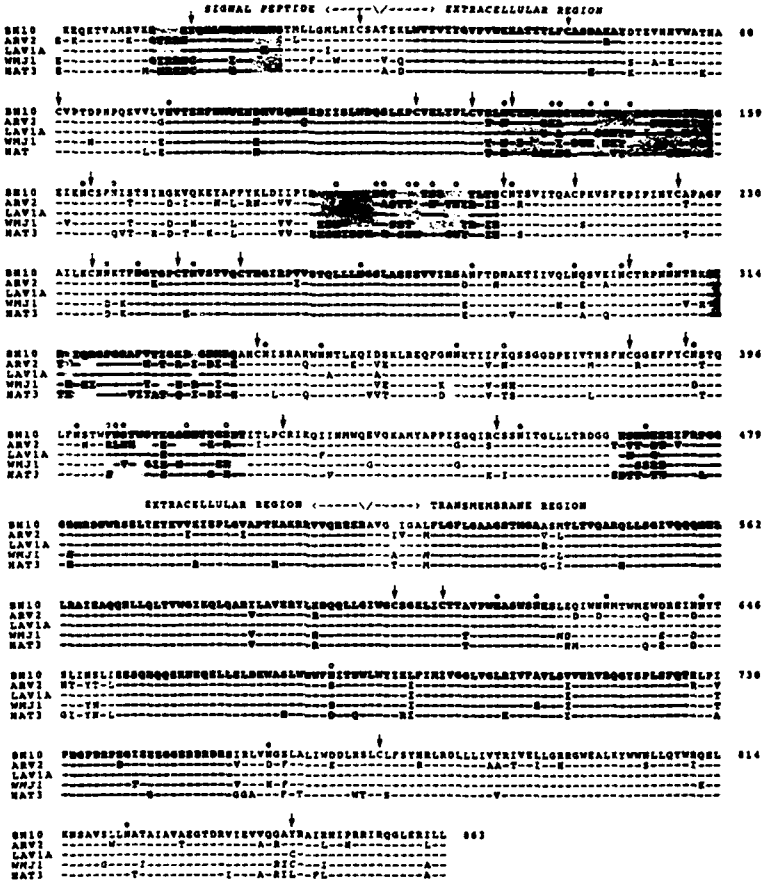


Figure 3. Amino Acid Sequences of the Entire Envelope Genes of Five Independent AIDS Virus Isolates. Sequence information for BH-10 (HTLV-IIIb), ARV-2, and LAV-1a was obtained from GENBANK and information for WMJ-1 and HAT-3 from the present study. Alignment of the sequences was performed pairwise with the assistance of PRTALN. Numbering of amino acids is from the first amino acid of BH-10 and regions corresponding to the signal peptide, extracellular envelope glycoprotein (gp 120), and transmembrane glycoprotein (gp 41) are shown. Dashes indicate amino acid identify with BH-10 and spaces indicate he absence of that amino acid. Arrows denote cysteine residues and solid and open circles denote conserved and nonconserved sites of potential N-linked glycosylation, respectively. Darkly shaded regions within the extracellular envelope glycoprotein correspond to regions of hypervariability as determined by both visual inspection and by computer analysis of variation (see Figure 4). Lightly shaded regions correspond to areas that are relatively highly conserved. Note that additional regions of intermediate variability are interspersed among highly variable and highly conserved areas.

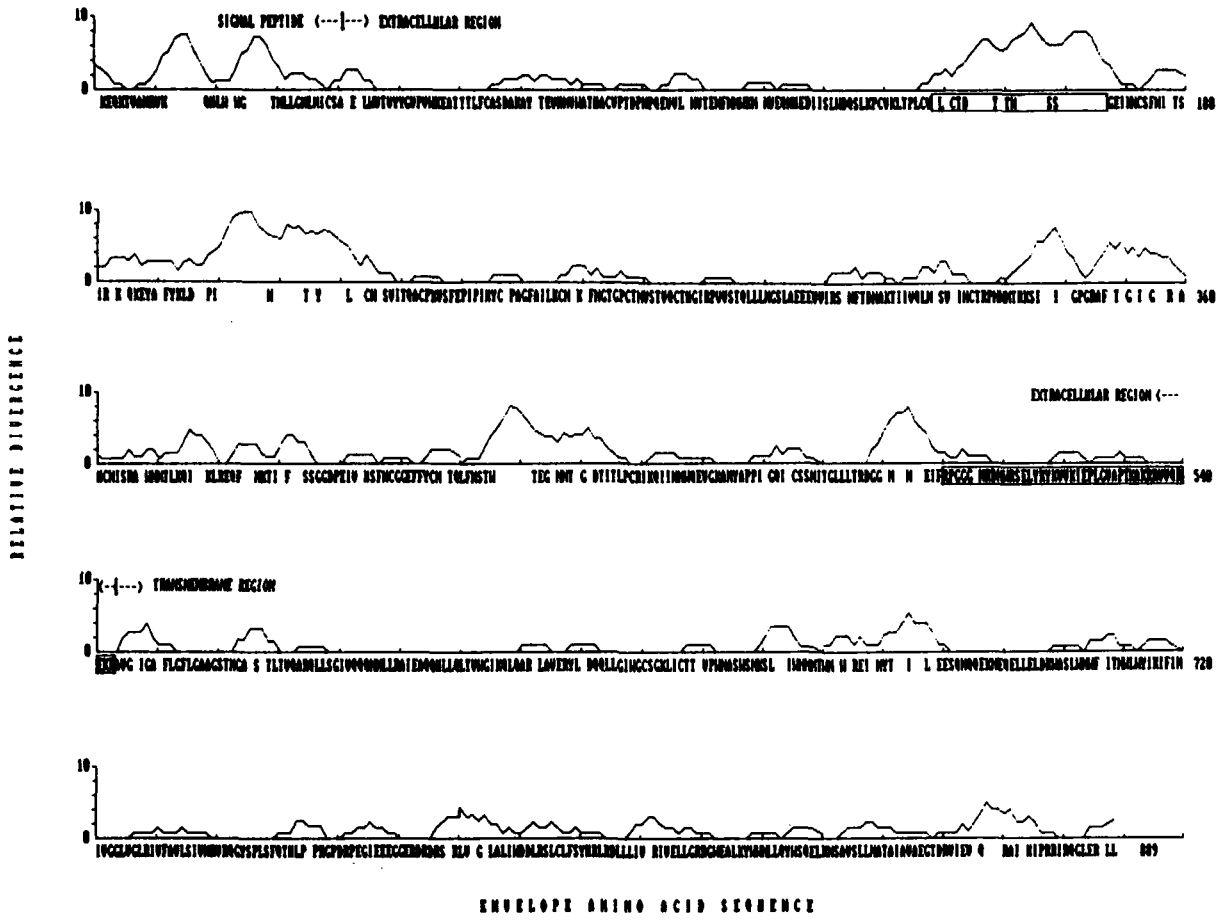


Figure 4. Computer Graphic Illustrating Relative Variation, or Divergence, in the Envelope Genes of Five Independent AIDS Virus Isolates, HTLV-IIIb (BH-10), LAV (LAV-1a) ARV (ARV-2), HTLV-III$_{RF}$ (HAT-3), and HTLV-III$_{WMJ-1}$ (WMJ-1). Relative degrees of variation were calculated and plotted from 0 (minimum divergence) to 10 (maximum divergence). Beneath the abscissa are single-letter codes for conserved amino acids that were present in five out of five or four out of five of the sequences. Numbering of amino acids is based on the envelope sequence as shown on the abscissa. The boxed amino acid sequences in line 180 and line 540 correspond to the variable and conserved regions whose secondary structures are depicted in Figure 5.

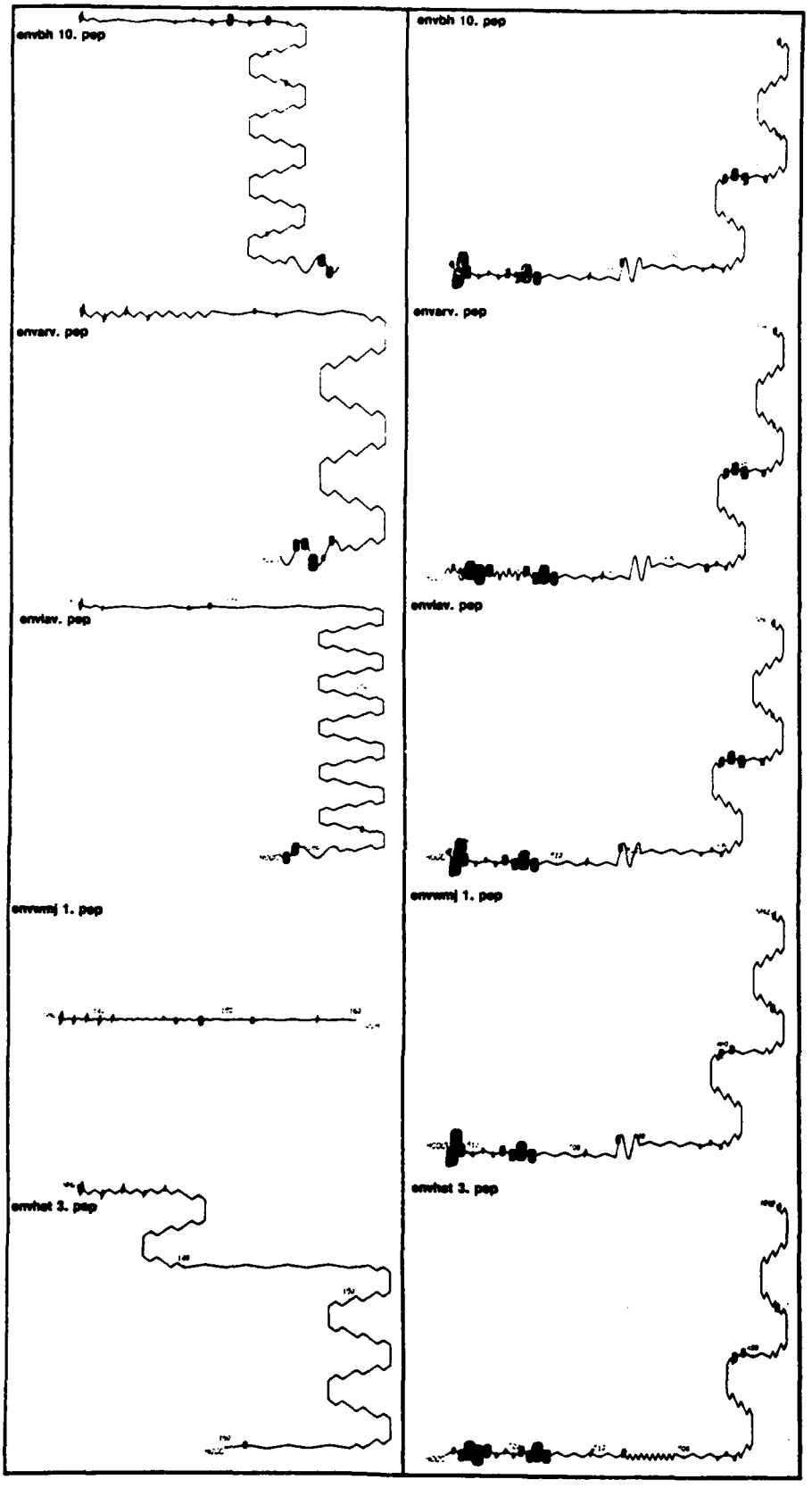
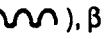
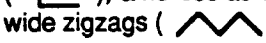

envbh 10. pep
envarv. pep
envlav. pep
envwmj 1. pep
envhat 3. pep
envbh 10. pep
envarv. pep
envlav. pep
envwmj 1. pep
envhat 3. pep

Figure 5

Figure 5. Predicted Secondary Structure and Hydrophilicity Patterns of Representative Variable and Conserved Envelope Regions of BH-10, ARV-2, LAV-1a, WMJ-1, and HAT-3.

The left hand panel corresponds to the first major variable region of the extracellular viral envelope, amino acids 137-158 in Figure 3. The right-hand panel corresponds to the largest conserved region, which, in addition, possesses structural and hydrophilicity properties suggestive of antigenicity. This region corresponds to amino acids 473-518 immediately 5' to the envelope processing site shown in Figure 3 . Ovals denote hydrophilic regions and diamonds hydrophobic regions. The radius of an oval or a diamond over a residue is proportional to the mean hydrophilicity or hydrophobicity as calculated for that residue plus the next four residues. Positions of β turns are indicated by turns of the chain (), α helices as bold coils (), β sheets as narrow zigzags (), and random coils as wide zigzags ().

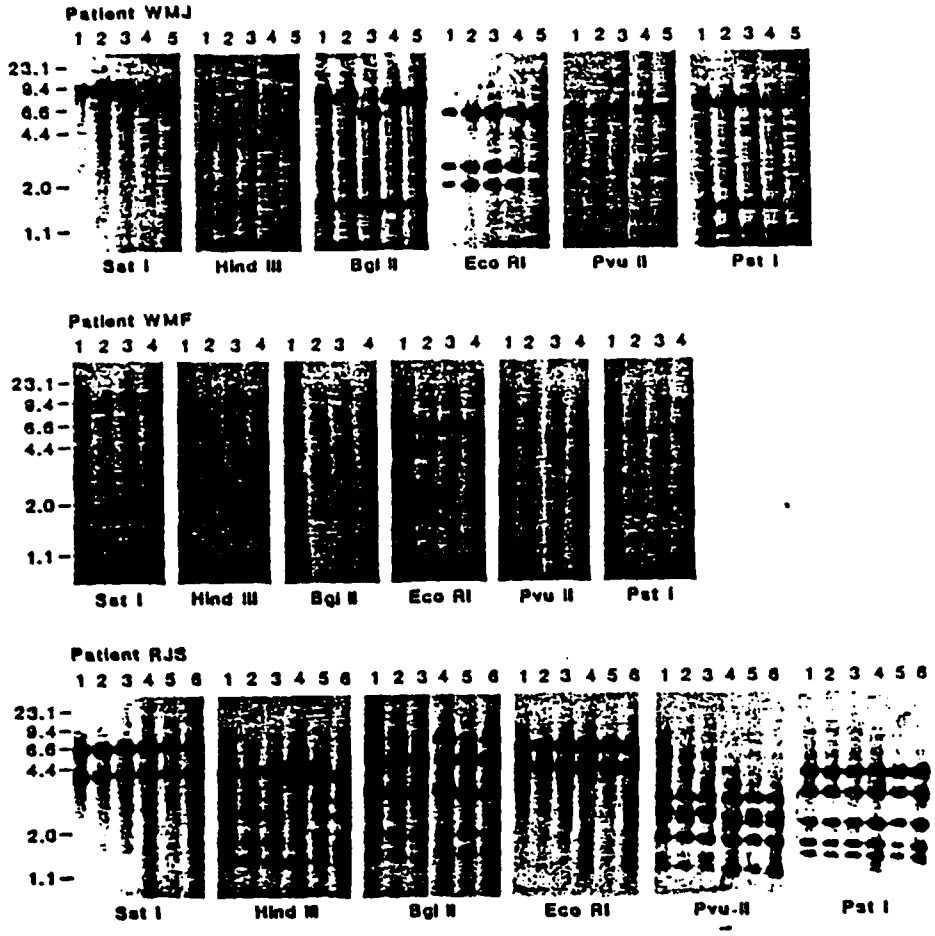


Figure 6. Southern blot hybridization analysis of sequential HIV-1 isolates. Virus was isolated from patients' peripheral blood mononuclear cells and propagated in tissue culture. High molecular weight DNA was prepared and Southern transfers performed. WMJ isolates 1, 2, 3, 4, and 5 were obtained from blood samples at intervals of 3, 4, 1, and 4 months, respectively. The last of these samples was obtained 1 month before the patient's death at 3 years of age. WMF isolates 1, 2, 3, and 4 were obtained at intervals of 6, 10, and 6 months, respectively, and RJS isolates 1, 2, 3, 4, 5, and 6 at intervals of 3, 4, 5, 7, and 4 months, respectively. Neither WMF nor RJS had developed AIDS. For each set of isolates from each patient, a panel of six restriction enzymes was used to map the viral genomes. The number of each virus isolate is indicated above the panels and the restriction enzymes are shown below each panel. Of note, the restriction enzymes Hind III, Bgl II, Eco RI, Pvu II, and Pst I were used in combination with Sst I, an enzyme known to cut the viral long terminal repeat. Thus, all viral bands generated represent internal viral fragments. Viral DNA's of patient RJS were also digested with the five restriction enzymes alone (without the addition of Sst I) because of the extra Sst I site present in the 3' half of all RJS isolates that was not present in WMJ or WMF isolates.

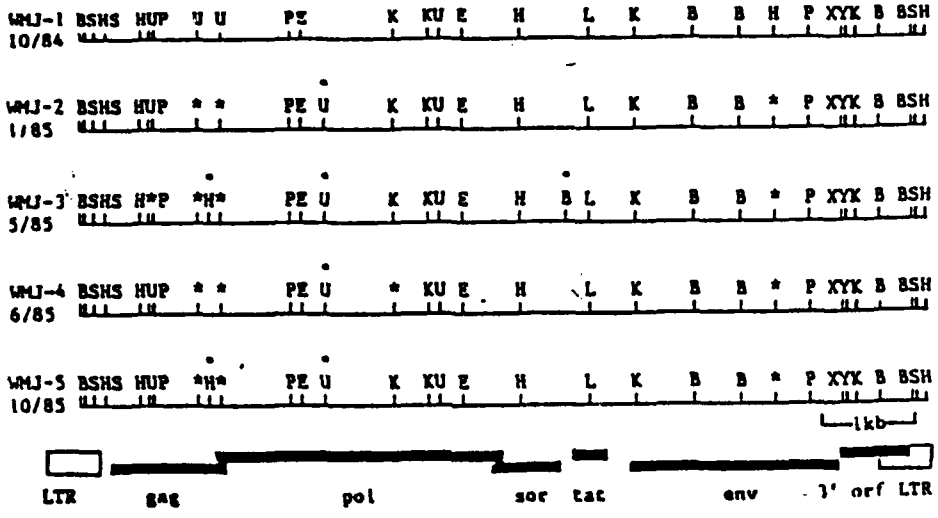


Figure 7. Restriction enzyme cleavage patterns of five sequential HIV-1 isolates from patient WMJ. Virus-infected DNA from WMJ-1 through WMJ-5 (see Fig. 6) was digested with the indicated restriction enzymes, Southern-blotted, and hybridized to subgenomic probes of HIV-1 (clones λBH5 and λBH8). For isolates WMJ-1, WMJ-2, and WMJ-3, the proviral DNA was molecularly cloned in λ phage and the restriction cleavage patterns confirmed. These clones, designated λWMJ-1 (V), λWMJ-2 (III), and λWMJ-3 (III), were subsequently subjected to nucleotide sequence determination (see Fig. 8). Dates of the phlebotomies from which the viruses were derived are indicated. Restriction sites are depicted by letters (B, Bgl II; S, Sst I; H, Hind III; U, Pvu II; P, Pst I; E, Eco RI; K, Kpn I; L, Sal I; X, Xho I; Y, Xba I). Restriction enzyme site differences among the five isolates are shown. Asterisks denote sites present in WMJ-1 but missing in other viruses, whereas solid dots denote sites that are absent in WMJ-1 but present in other viruses. Note that the restriction enzyme differences among the five viruses are not consistent with a direct, sequential evolution of changes from isolates WMJ-1 to WMJ-2, and so on, to WMJ-5 (see text). Beneath the restriction maps is drawn to scale the genomic organization of HIV-1.

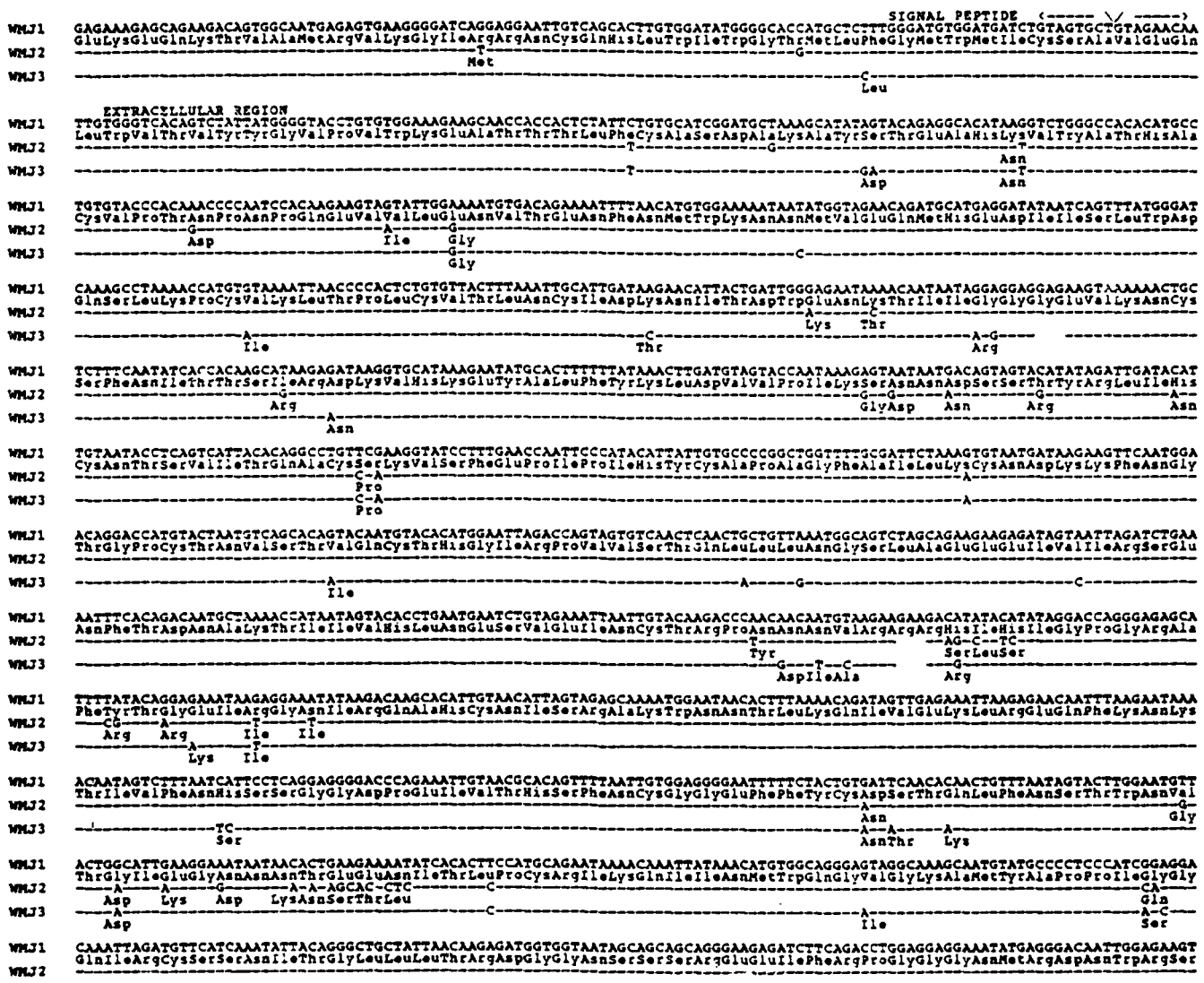
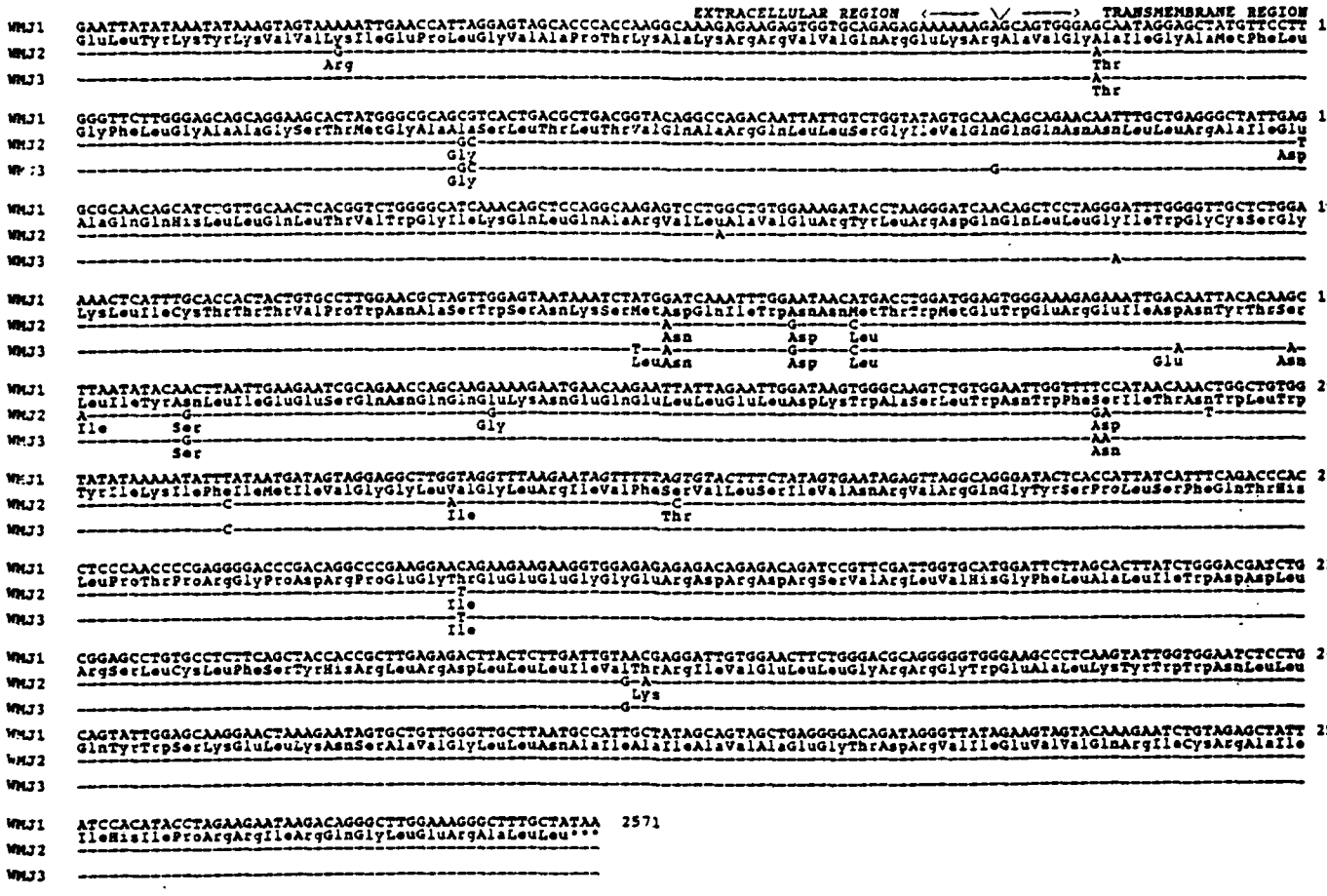


Figure 8. Nucleotide and amino acid sequence alignment of the entire envelope genes of three sequential *virus isolates from patient WMJ.* Molecular clones of WMJ-1, WMJ-2, and WMJ-3 [designated λWMJ-1 (V), λWMJ-2 (III), and λWMJ-3 (III), respectively] were obtained and their envelope genes sequenced. Alignment of the sequences was performed pairwise with the assistance of NUCALN. Sequences are numbered according to WMJ-1 starting with the first nucleotide following a TAA stop codon. Genomic regions corresponding to the signal peptide, the extracellular glycoprotein (gp120), and the membrane-associated glycoprotein (gp41) are shown. Dashes indicate nucleotide identify and spaces indicate deletions in WMJ-2 and WMJ-3 with respect to WMJ-1.

Figure 8 (cont.)



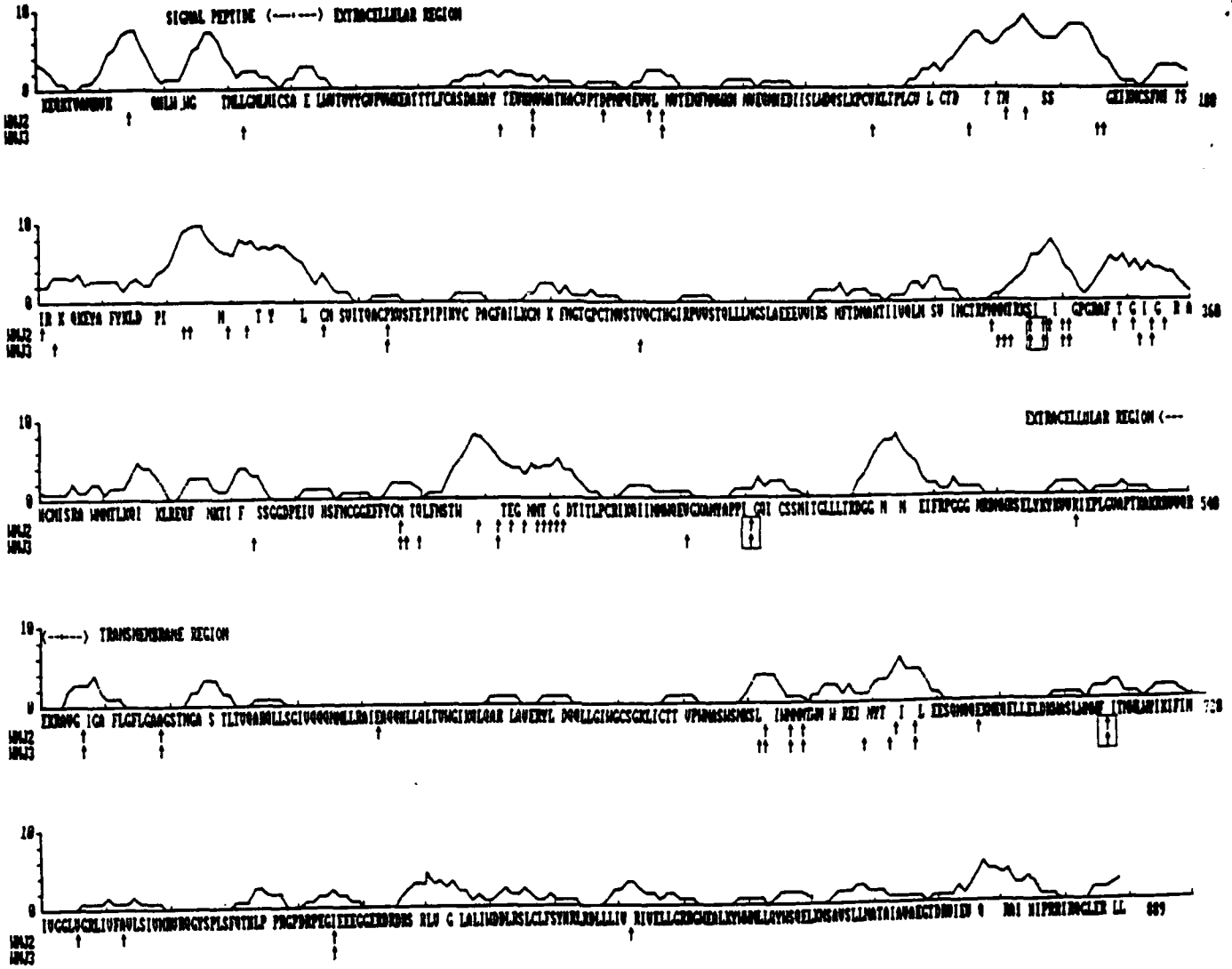
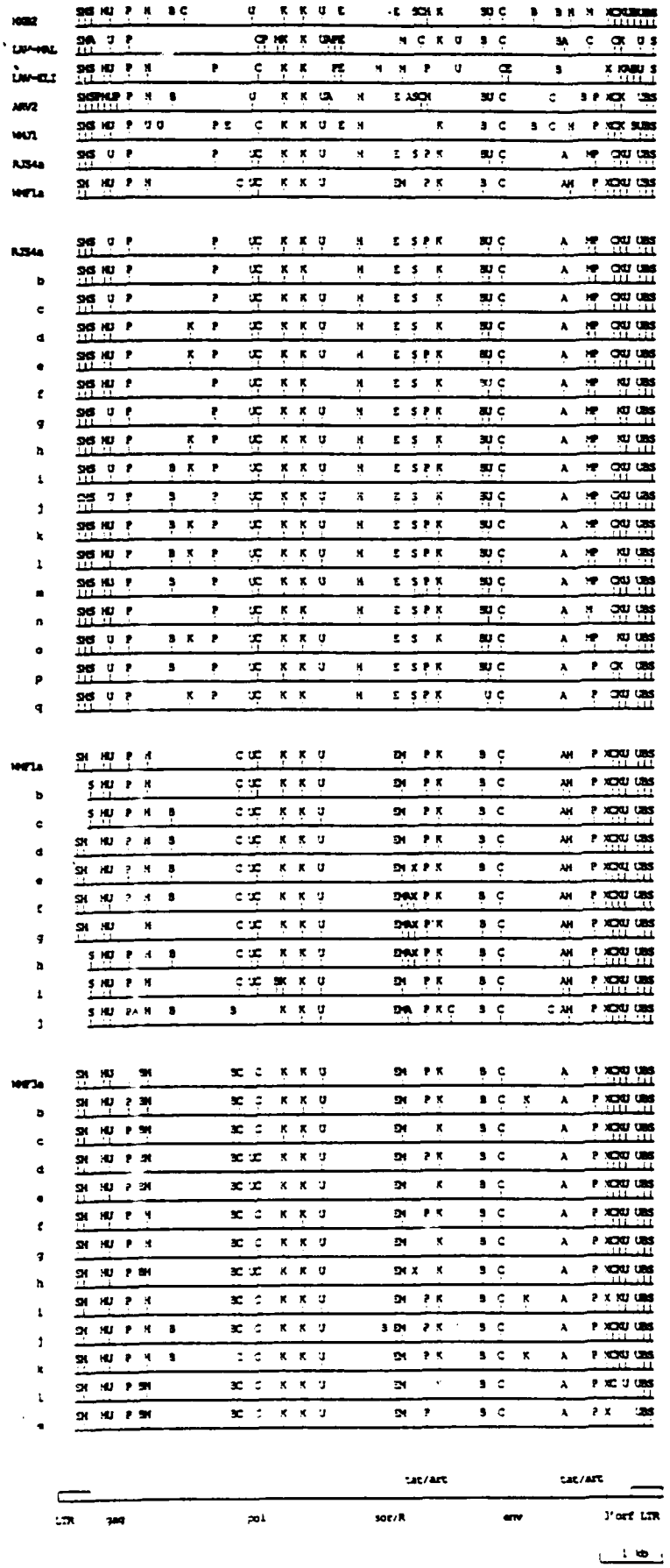


Figure 9. Computer graphic illustrating relative variation, or divergence, in the envelope genes of five independent AIDS virus isolates and the position of amino acid differences among WMJ-1, WMJ-2, and WMJ-3. The five independent isolates were HTLV-IIIb (BH-10), LAV (LAV-1a), ARV (ARV-2), HTLV-III_{RF} (HAT-3), and HTLV-III_{WMJ-1} (WMJ-1). Their relative degrees of variation were calculated and plotted on the ordinate from 0 (minimum divergence) to 10 (maximum divergence). Beneath the abscissa are single letter codes for conserved amino acids as defined by identity in five out of five or four out of five of the sequences. Arrows beneath this relative consensus sequence indicate the position of amino acid changes in WMJ-2 and WMJ-3 compared to WMJ-1. Double arrows represent identical amino acid substitutions in WMJ-2 and WMJ-3 compared to WMJ-1 while boxed double arrows indicate different amino acid substitutions Numbering of amino acids is based on the envelope sequence as shown on the abscissa. Note that most of the amino acid differences among the three WMJ isolates are confined to regions of either high or intermediate envelope variability.



Virus	Pattern	
	E S X B U H P K M C A	
HXB2	1 1 1 1 1 1 1 : 1 1 1 1	
LAV-MAL	2 2 2 2 2 2 2 : 2 2 2	
LAV-ELI	3 3 1 3 3 3 3 2 3 3 1	
ARV2	4 1 1 4 4 4 4 2 4 4 4	
WAU1	5 3 1 5 5 5 5 : 4 5 1	
RJS4a	4 1 2 6 : 6 6 1 1 6 5	
HNFLa	4 4 1 6 6 7 7 : 5 7 5	

Virus	Pattern	Clone	Total
RJS4	E S X B U H P K M C A		Number
a	1 1 1 1 1 1 1 1 1 1 1	1, 11, 16, 39, 46	5
b	1 1 1 1 2 2 2 1 1 1 1	2, 23, 26	3
c	1 1 1 1 1 1 2 1 1 1 1	18, 34	2
d	1 1 1 1 1 2 2 2 1 1 1	15, 40	2
e	1 1 1 1 1 2 1 2 1 1 1	20, 24	2
f	1 1 1 1 2 2 2 1 1 2 1	12, 31	2
g	1 1 1 1 1 1 1 1 1 2 1	7	1
h	1 1 1 1 1 2 2 2 1 2 1	32	1
i	1 1 1 2 1 1 1 2 1 1 1	42	1
j	1 1 1 2 1 1 2 1 1 1 1	6	1
k	1 1 1 2 1 2 1 2 1 1 1	22	1
l	1 1 1 2 1 2 1 2 1 2 1	14	1
m	1 1 1 2 1 2 1 1 1 1 1	21	1
n	1 1 1 1 2 2 3 1 1 1 1	45	1
o	1 1 1 2 1 3 2 2 1 2 1	19	1
p	1 1 1 2 3 1 1 1 2 1 1	43	1
q	1 1 1 3 2 1 1 2 2 1 1	4	1

Virus	Pattern	Clone	Total
HNFl	E S X B U H P K M C A		Number
a	1 1 1 1 1 1 1 1 1 1 1	5, 9, 15, 16, 18	5
b	1 2 1 1 1 1 1 1 1 1 1	4, 14, 19	3
c	1 2 1 2 1 1 1 1 1 1 1	11, 13	2
d	1 1 1 2 1 1 1 1 1 1 1	10	1
e	1 1 2 2 1 1 1 1 1 1 1	1	1
f	1 1 2 2 1 1 1 1 1 1 2	12	1
g	1 1 2 1 1 1 2 1 1 1 2	2	1
h	1 2 2 2 1 1 1 1 1 1 2	6	1
i	1 2 1 3 1 1 1 1 1 1 1	3	1
j	1 2 1 4 2 1 1 1 1 2 2	17	1

Virus	Pattern	Clone	Total
HNFJ	E S X B U H P K M C A		Number
a	1 1 1 1 1 1 1 1 1 1 1	1, 110, 111	3
b	1 1 1 1 1 1 1 2 1 1 1	103, 107	2
c	1 1 1 1 1 1 2 1 1 1 1	3, 108	2
d	1 1 1 1 2 1 1 1 1 1 1	10, 104	2
e	1 1 1 1 2 1 2 1 1 1 1	102	1
f	1 1 1 2 1 1 1 1 1 1 1	11	1
g	1 1 1 2 1 1 2 1 1 1 1	101	1
h	1 1 2 1 2 1 2 1 1 1 1	106	1
i	1 1 1 2 1 1 1 2 1 2 1	9	1
j	1 1 1 3 1 1 1 1 1 1 1	7	1
k	1 1 1 4 1 1 1 2 1 1 1	12	1
l	1 1 1 1 1 1 2 3 1 1 1	105	1
m	1 1 1 1 3 1 1 4 1 2 1	109	1

Figure 10

Figure 10. Restriction endonuclease cleavage patterns for 7 independent HIV-1 viral clones (HXB2; LAV-Mal; LAV-Eli; ARV2; WMJ1; RJS4a; WMF1a) and for 17 distinguishable clones from isolate RJS4, 10 different clones from isolate WMF1, and 13 distinct clones from isolate WMF3. To the right of each set of restriction maps is a table summarizing the different restriction enzyme patterns for each representative clone within that set (E-EcoRI; S-SstI; X-XhoI; B-BglII; U-PvuII; H-HindIII; P-PstI; K-KpnI; M-BamHI; C-CvnI; A-XbaI). The top clone pattern in each set is arbitrarily designated 11111111111 and differences in subsequent clone patterns are identified sequentially. The laboratory clone (e.g., RJS4.1; RJS4.11; RJS4.16; etc.) designation and the total number of viral clones represented by each pattern are also indicated.

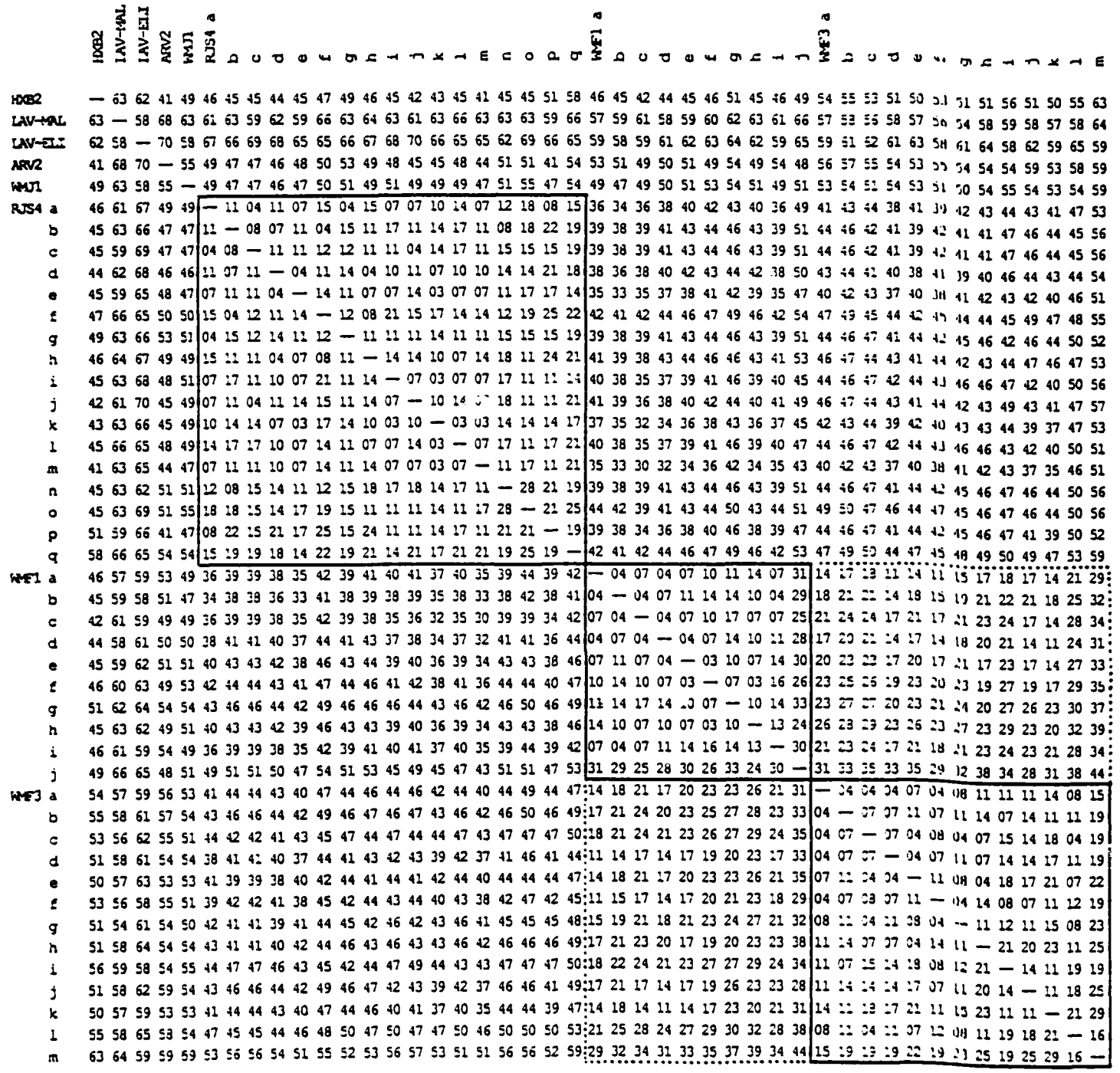


Figure 11

Figure 11. Analysis of variation among viral DNA clones of HIV-1. Each of the 45 different viral clone patterns depicted in Figure 10 was compared pairwise to every other clone pattern and the percentage difference in restriction endonuclease cleavage sites between each pair was calculated as follows:

$$\frac{A + B}{C} \times 100 = \% \text{ Restriction Site Differences}$$

where, A equals the number of restriction sites present in one clone (X) that are missing in the other clone (Y); B equals the number of restriction sites present in clone Y that are missing in clone X; C equals the total number of restriction sites present in clones X and Y combined, with identical sites in the pair counted only once. For example, if clone X and clone Y had 25 restriction sites in common and no additional sites, they would have:

$$\frac{0 + 0}{25} \times 100 = 0\% \text{ Restriction Site Differences}$$

If clones X and Y each had 25 restriction sites but none were in common, they would have:

$$\frac{25 + 25}{50} \times 100 = 100\% \text{ Restriction Site Differences}$$

This analytic approach allowed each of the 45 distinct clone patterns to be compared with every other clone. Because EcoRI and SstI enzymes were used to clone the RJS and WMF viruses, respectively, these enzyme sites were not included in the analysis of variation among the RJS4, WMF1, and WMF3 clones. Because the two terminal SstI sites in all clones except WMF1b, c, h, i, and j actually represent redundant sequences in their LTRs, only one of the two was included in the analysis of variation. Thus, the first pair of clones compared in Table 1 (HXB2 versus LAV-Mal) differ by a total of 27 out of 43 restriction sites, or 63%.

$$\frac{17 + 10}{43} \times 100 = 63\% \text{ Restriction Site Differences}$$

Boxed in solid lines are the difference scores for viral clones derived from the RJS4, WMF1, and WMF3 isolates. The dashed lines highlight the comparison of viral clones between isolates WMF1 and WMF3, both of which were derived from the same chronically infected individual 16 months apart. In order to compare statistically the extent of dissimilarity among clones from within individual isolates (RJS4a-q; WMF1a-j; WMF3a-m) versus clones from independent (unrelated) isolates (HXB2, LAV-Mal, LAV-Eli, ARV2, WMJ1), the individual data points, not mean percentages, of the respective groups were compared by chi square analysis.

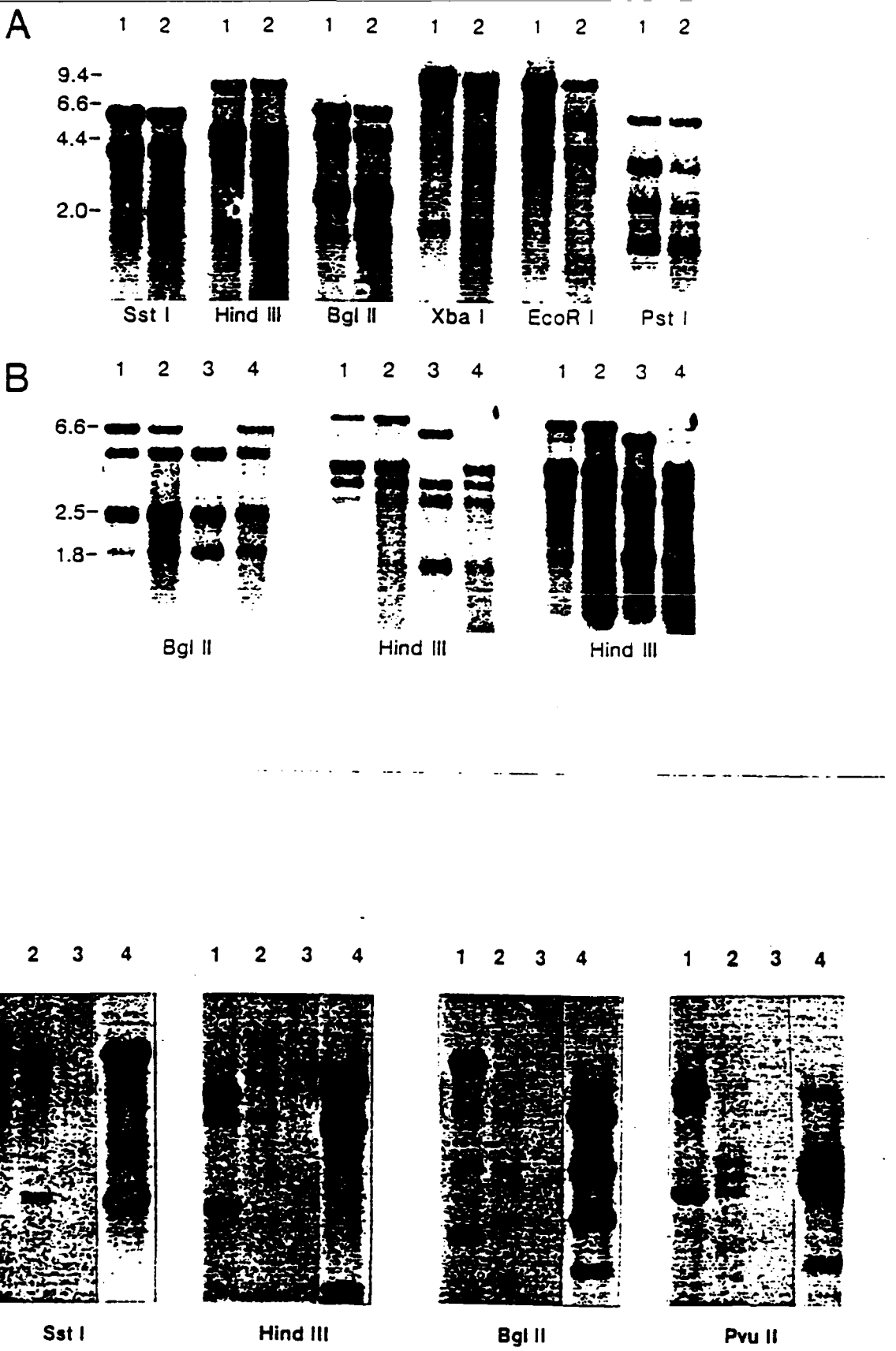

A
1 2 1 2 1 2 1 2 1 2 1 2
9.4-
6.6-
4.4-
2.0-
Sst I
Hind III
Bgl II
Xba I
EcoR I
Pst I
B
1 2 3 4 1 2 3 4 1 2 3 4
6.6-
2.5-
1.8-
Bgl II
Hind III
Hind III
C
1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
9.4—
6.6—
4.4—
2.3—
1.1—
Sst I
Hind III
Bgl II
Pvu II

Figure 12

Figure 12. Southern blot restriction enzyme cleavage patterns of parallel and sequential isolates of HIV-1 (Panels A and B) and of cultured versus uncultured virus (Panel C). Panel A, lanes 1 and 2 represent viral isolates obtained by co-cultivating a single (split) sample of PBMC from an HIV-1 infected subject (RJS) with PHA-stimulated lymphocytes from two different, HLA-unrelated normal donors. Panel B, lanes 1 and 2, are the same as lanes 1 and 2 from panel A. Lanes 3 and 4 correspond to virus isolates derived from the PBMC of this same HIV-1 infected subject taken 2 weeks and 4 weeks after the first phlebotomy and co-cultured again with PBMC from two other unrelated normal donors. The right-hand blot in panel B is identical to the center blot except that the exposure time was lengthened from two to five days in order to identify faint signals. Notice that the predominant HindIII patterns of isolates 3 and 4 are distinguishable from isolates 1 and 2, yet with long exposure films persistence of common viral genotypes can be identified in all isolates. Panel C, Southern blot-hybridization patterns of HIV-1 DNA extracted directly from uncultured brain tissue (lanes 2) and from HIV-1 viral cultures derived by lymphocyte co-cultivation with the same brain tissue specimen (lanes 4). Lanes 1 show an HIV-1 positive control DNA from an unrelated virus isolate and lanes 3 represent uninfected negative control DNA.

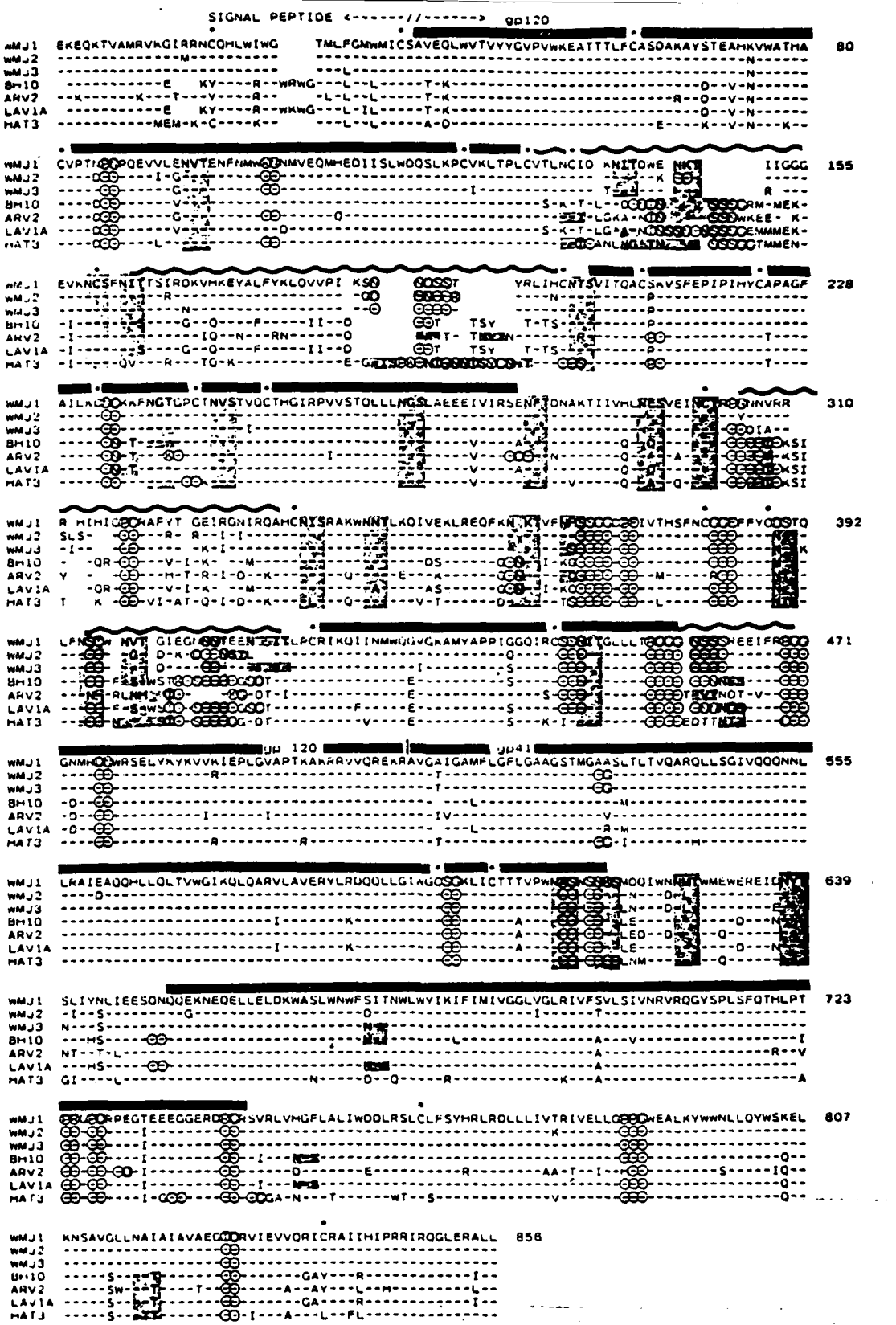


Figure 13. Amino acid sequences of the entire envelope genes of seven HIV isolates, WMJ1, WMJ2, and WMJ3 are sequential isolates from the same patient: BH10, ARV2, LAV1-a, and HAT-3 represent independent virus isolates. Alignment of the sequences was done with the assistance of PRTLAN. Numbering of amino acid sequences is done from the first residue of the envelope protein open reading frame of WMJ1 up to the end. Dashes indicate amino acid identity with WMJ1, spaces indicate the absence of that amino acid. N-glycosylation sites are indicated as shaded areas, amino acids in β turn regions are encircled. Symbols: *, cysteine residues; —, conserved regions; ~~~, variable regions.

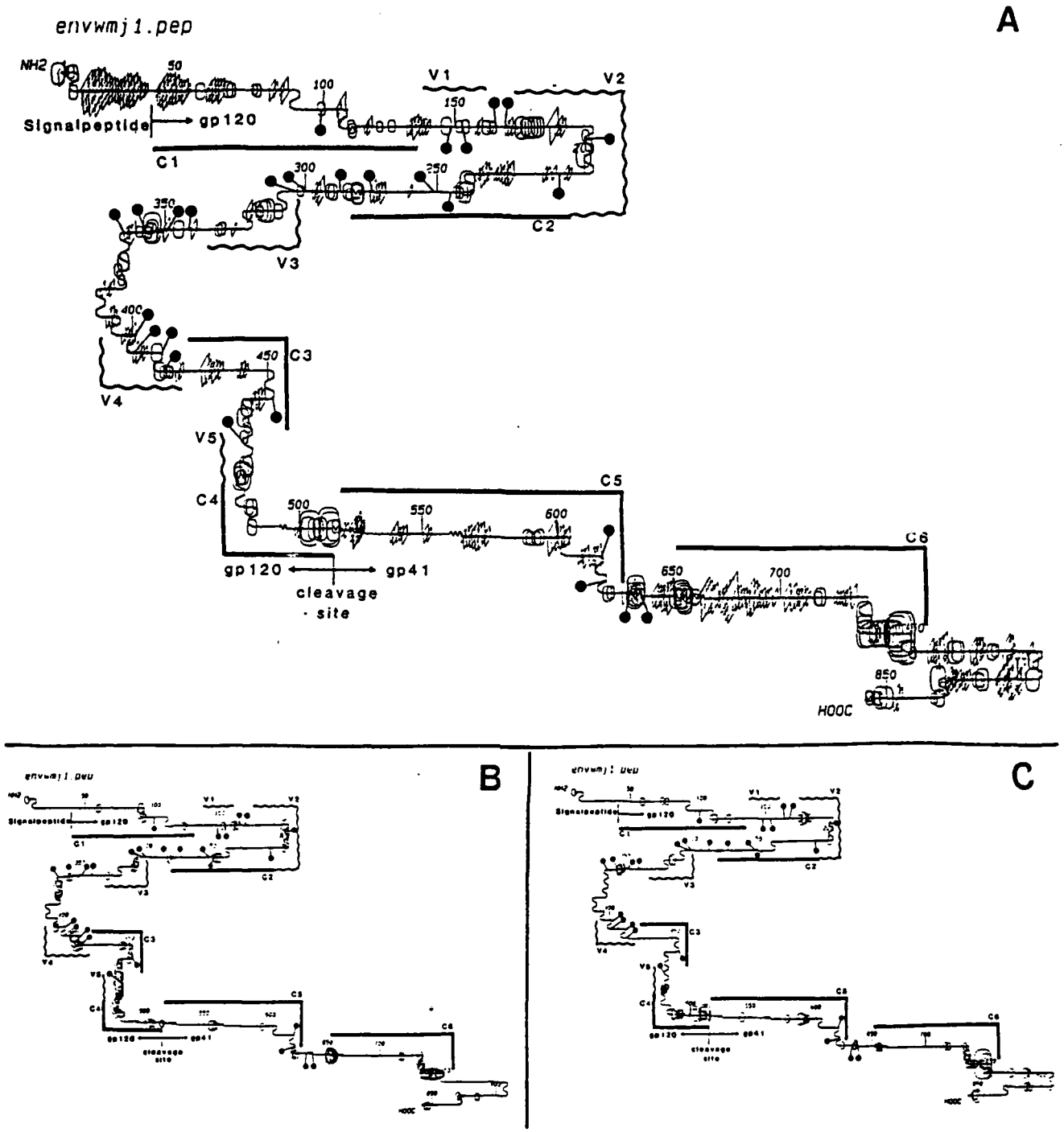


Figure 14. Chou-Fasman prediction of the envelope protein gp160 derived from the sequence of WMJ1. The computer plots start at the methionine at residue 9 of the open reading frame. The probability of the occurrence of α-helices (~~~), β-pleated sheets (⋀⋀), random coils (~~~), and β turn regions (⊂⊃), were evaluated by using stringent conditions. The parameters for hydrophilicity, flexibility, and surface probability were averaged over five amino acid residues, with a limit of 0.7 for hydrophilicity, 1.040 for flexibility, and 5.0 for surface probability. Symbols: ●—, N-glycosylation sites; ⎯, conserved regions; variable regions. (A) Secondary structures superimposed with the values for hydrophilicity; ◊, hydrophilic regions, ◆, hydrophobic regions. (B) Secondary structures superimposed with the values for flexibility; ◊, flexible regions. (C) Secondary structures superimposed with the values for surface probability; ◊, high surface probability.

Region	Length	Amino acid deletions/ insertions	Conserved amino acids		Glycosylation sites	Conserved glycosylation sites		β-Turns	Conserved β-turns	
			No.	%		No.	%		No.	%
C1	97/97/97	0	84	87	1/1/1	1	100	4/4/4	2	50
38–134	97/97/97/97				1/1/1/1			2/4/2/4		
V1	20/20/20	0–15	3	10	2/2/2	0	0	0/2/0	0	0
135–154	24/23/29/25				1/2/3/3			8/5/11/7		
	8/8/8	0	7	87	1/1/1	1	100	0/0/0	0	100
155–162	8/8/8/8				1/1/1/1			0/0/0/0		
V2	41/41/41	0–18	11	21	2/3/2	1	20	5/6/5	0	0
163–203	39/44/39/52				2/4/3/5			2/2/2/17		
C2	76/76/76	0	68	89	3/3/3	2	50	2/2/2	2	50
204–279	76/76/76/76				4/3/4/3			2/4/2/4		
	25/25/25	0	19	76	3/3/3	3	100	0/0/0	0	0
280–304	25/25/25/25				3/3/3/3			0/3/0/0		
V3	26/26/26	0–5	7	25	0/0/0	0	0	4/2/4	2	25
305–330	27/26/27/26				1/1/1/1			8/6/8/8		
	65/65/65	0–1	48	72	5/6/6	4	67	11/9/10	8	72
331–395	66/66/66/65				5/5/5/6			11/10/11/10		
V4	19/19/19	0–11	4	17	3/2/3	0	0	4/6/4	0	0
396–414	23/18/23/19				2/1/2/3			11/4/11/9		
C3	44/44/44	0	38	86	1/1/1	1	100	7/7/7	4	57
415–458	44/44/44/44				1/1/1/1			7/7/7/4		
V5	11/11/11	0–2	2	16	1/1/1	0	0	5/5/5	1	18
459–469	11/12/11/12				1/1/1/1			5/1/6/1		
C4	41/41/41	0	36	87	0/0/0	0	100	4/4/4	4	100
470–510	41/41/41/41				0/0/0/0			4/4/4/4		
C5	106/106/106	0–1	95	90	2/2/2	2	100	6/8/8	6	75
511–616	105/106/105/106				2/2/2/2			6/6/6/8		
	38/38/38	0	17	44	2/2/2	2	100	2/0/0	0	0
616–653	38/38/38/38				2/2/2/2			2/0/2/2		
C6	92/92/92	0	84	85	0/0/1	0	0	6/6/6	6	66
654–745	92/92/92/92				1/0/1/0			6/8/6/10		
	111/111/111	0	77	69	0/0/0	0	0	5/5/5	4	57
746–856	111/111/11/111				2/1/2/2			5/4/5/7		

Region	Length	Hydrophilic β-turns	Conserved hydrophilic β-turns		Hydrophilicity		Surface probability		Flexibility	
			No.	%	No.	%	No.	%	No.	%
C1	97/97/97	2/2/2	2	100	10/8/7	7–10	7/7/9	7–10	11/1/11	8–10
38–134	97/97/97/97	2/2/2/2			8/9/9/9		8/9/7/10		11/11/8/12	
V1	20/20/20	0/0/0	0	0	5/5/7	13–33	0/0/0	0–13	6/7/6	28–82
135–154	24/23/29/25	3/0/1/0			9/9/4/4		2/3/3/0		14/19/17/14	
	8/8/8	0/0/0	0	100	0/0/0	0	0/0/0	0	0/0/0	0
	8/8/8/8	0/0/0/0			0/0/0/0		0/0/0/0		0/0/0/0	
V2	41/41/41	4/4/4	0	0	9/17/6	14–41	9/12/4	10–32	8/11/12	20–46
163–203	39/44/39/52	2/0/2/3			7/9/7/9		5/5/5/12		11/9/11/24	
C2	76/76/76	2/2/2	2	67	9/8/9	7–11	0/0/0	0	5/5/5	7–11
204–279	76/76/76/76	2/3/2/2			5/5/5/9		0/0/0/0		7/8/9/6	
	25/25/25	0/0/0	0	0	5/5/5	8–20	0/0/0	0	2/2/2	8–12
280–304	25/25/25/25	0/2/0/0			2/2/2/4		0/0/0/0		0/3/0/2	
V3	26/26/26	3/1/3	1	16	7/5/8	19–34	2/5/5	7–22	5/0/3	0–57
305–330	27/36/27/26	6/7/6/4			7/7/7/9		6/2/6/3		12/8/12/15	
	65/65/65	4/4/4	4	57	11/11/11	12–17	5/5/5	6–8	8/8/9	12–25
331–395	66/66/66/65	6/7/6/4			11/8/9/10		5/5/4/4		17/10/10/10	
V4	19/19/19	1/4/1	0	0	5/7/7	2–11	0/0/0	0–6	9/13/9	10–20
396–414	23/18/23/19	8/2/6/3			3/4/3/1		0/1/0/0		12/7/12/13	
C3	44/44/44	4/4/4	4	100	0/1/0	0–11	0/0/0	0	5/5/5	5–11
415–458	44/44/44/44	4/4/4/4			2/3/5/2		0/0/0/0		5/2/5/4	
V5	11/11/11	2/2/2	0	0	10/10/10	1–90	4/5/5	0–45	10/6/10	54–100
459–469	11/12/11/12	1/0/1/0			8/2/1/7		0/0/0/1		11/9/9/11	
C4	41/41/41	4/4/4	4	100	12/12/12	29–32	18/19/19	43–46	12/11/12	24–29
470–510	41/41/41/41	4/4/4/4			13/13/13/12		19/19/19/19		12/12/12/10	
C5	106/106/106	0/0/0	0	100	5/5/5	5	8/8/8	5–7	5/5/5	6
511–616	105/106/105/106	0/0/0/0			5/5/5/5		8/8/8/5		6/5/6/5	
	38/38/38	0/0/0	0	0	7/5/5	13–23	4/4/4	10–13	3/3/3	7–10
616–653	38/38/38/38	2/0/2/2			7/5/7/4		3/4/3/4		3/3/3/2	
C6	92/92/92	6/6/6	6	67	34/34/34	34	29/20/23	20–29	35/35/36	32–40
654–745	92/92/92/92	6/8/6/10			34/34/34/34		19/22/24/28		33/32/37/40	
	111/111/111	5/5/5	2	39	20/20/19	15–18	13/13/13	10–12	9/9/9	7–10
746–856	111/111/111/111	5/2/5/7			20/17/20/20		12/12/12/11		10/8/10/11	

*Values are arranged in the following mode: WMJ1/ WMJ2/WMJ3
BH10/ ARV2/ LAV1A/HAT3.

Figure 15

Figure 15. Parameters and values for conserved and variable regions, calculated from the sequences of seven HIV envelope proteins (gp 160) (see Figure 13).

Epitope no.	Region	Amino acid residues in virus strain:						
		WMJ1	WMJ2	WMJ3	BH10	ARV2	LAV1A	HAT3
I	V1	137–154	137–154	137–153	137–157	137–158	137–163	137–158
II	V2	186–203	186–203	185–202	188–203	189–209	195–209	189–217
III	C2	232–246	232–246	228–245	232–246	240–253	240–253	249–261
IV	V3	300–320	300–319	299–318	300–321	300–327	308–328	316–355
V		358–375	357–374	356–373	360–377	367–384	367–384	375–391
VI	V4	394–412	393–411	392–410	397–418	404–420	404–425	410–427
VII	C3	445–458	444–457	443–456	451–464	453–466	457–470	459–472
VIII	V5	459–469	458–468	457–467	465–475	467–478	471–481	473–484
IX	C4	470–483	469–482	468–481	476–489	478–492	482–496	485–499
X		611–637	610–636	609–635	616–643	620–646	623–649	627–653
XI	C6	724–745	723–744	722–743	729–750	733–754	736–757	740–761

Figure 16. Predicted antigenic epitopes in seven strains of HIV-1.

HTLV-4/PK82 provirus λ PKE102	LTR	gag	pol	sor	X	R	tat	art	env	3' orf	total
compared with:											
HTLV-4/PK190 cDNA pV2	4/519 0.8%									3/736 0.4%	6/850 0.7%
STLV-3_{AGM}/K78 provirus λ K2-10	13/807 1.6%			5/227 2.2%	7/336 2.1%	2/291 0.7%	1/393 0.3%	0/324 0%	21/2643 0.8%	6/789 0.8%	48/4520 1.1%
STLV-3_{AGM}/K6W provirus	13/796 1.6%	7/637 1.1%		1/227 0.4%	—† —	—† —	1/352† 0.3%	0/324 0%	16/2643 0.6%	6/776 0.7%	44/5036 0.9%
STLV-3_{AGM}/K1 cDNA pBS3	1/668 0.1%								2/280 0.7%	3/780 0.4%	4/1160 0.4%
STLV-3_{AGM}/K6W provirus compared with:	30/780 3.9%	43/1518 2.8%	95/3168* 3.0%	26/642* 4.0%	(8/336)‡ (2.4%)	(16/291)‡ (5.5%)	27/352† 7.6%	19/324 5.9%	124/2643 4.7%	62/776 7.7%	351/9090 3.9%
SIV_{Mac}/Mm142-83 provirus SIV-1											

Figure 17. Nucleotide sequence differences between molecular clones of HTLV-4, STLV-3_{AGM} and STLV-3(SIV)$_{MAC}$

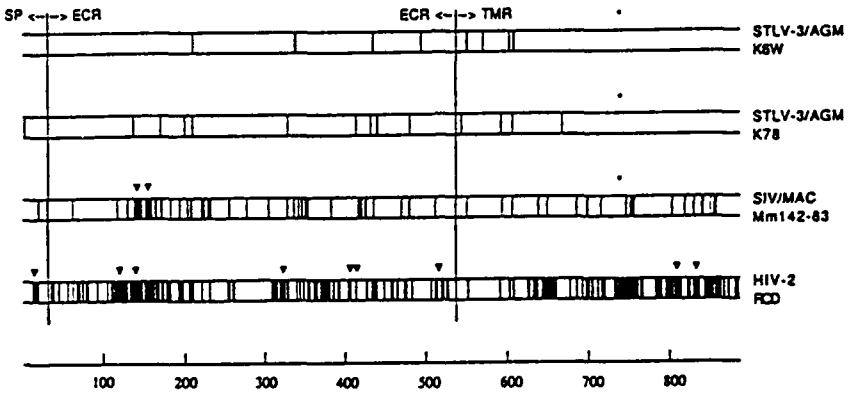


Figure 18. Distribution of amino-acid sequence differences in the envelope genes of STLV-3_{AGM}, STLV-3 $(SIV)_{MAC}$, and HIV-2 compared to HTLV-4. Vertical lines represent the position of single amino-acid sequence differences in the envelopes of STLV-3_{AGM}/K6W, STLV-3_{AGM}/K78, SIV_{MAC}/Mm142-83, and HIV-2_{ROD} compared with HTLV-4/PK82, as determined by pairwise alignment using the program PRTALN. Arrows depict the position of amino-acid insertions and deletions. An asterisk depicts the location of a TAG in-frame stop codon in the transmembrane region which is present in the clones of HTLV-4, STLV-3_{AGM} and STLV-3 $(SIV)_{MAC}$, but not in HIV-2_{ROD}. SP, signal peptide; ECR, extracellular region; TMR, transmembrane region. Relative amino-acid positions are shown on the bottom scale.

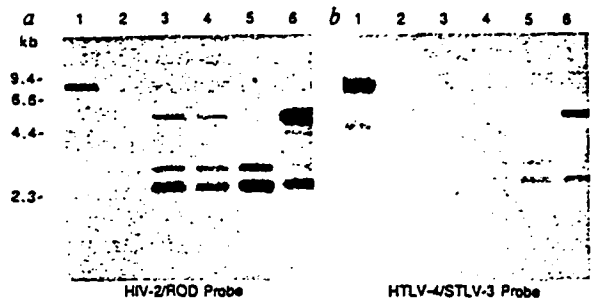


Figure 19. Determination of relative nucleotide sequence homology of a newly isolated Senegalese retrovirus (isolate ST) to HIV-2_{ROD} (panel *a*) and HTLV-4/STLV-3 (panel *b*). Panels *a* and *b* are identical except for the viral-specific nucleic acid probe used. Lane 1, HTLV-4/PK82; Lane 2, normal human DNA; Lanes 3-6, 4 different subcultures derived from the parental ST virus isolate. Preferential hybridization of the HIV-2 specific probe to ST isolate DNA is seen (panel *a*, lanes 3-6). The HTLV-4/STLV-3 specific probe hybridizes much less efficiently to ST DNA (panel *b*, lanes 3-6) but hybridizes to HTLV-4 DNA on the same filter (panel *b*, lane 1). Lanes 5 and 6 (panel *a*) show two distinct ST viral genotypes polymorphic in a single *Bam*HI restriction site. Virus cultures (lanes 3 and 4) contain a mixture of these two predominant viral forms which also differ in their *Xba*I restriction patterns (data not shown). Size markers are given on the left in kilobases (kb)

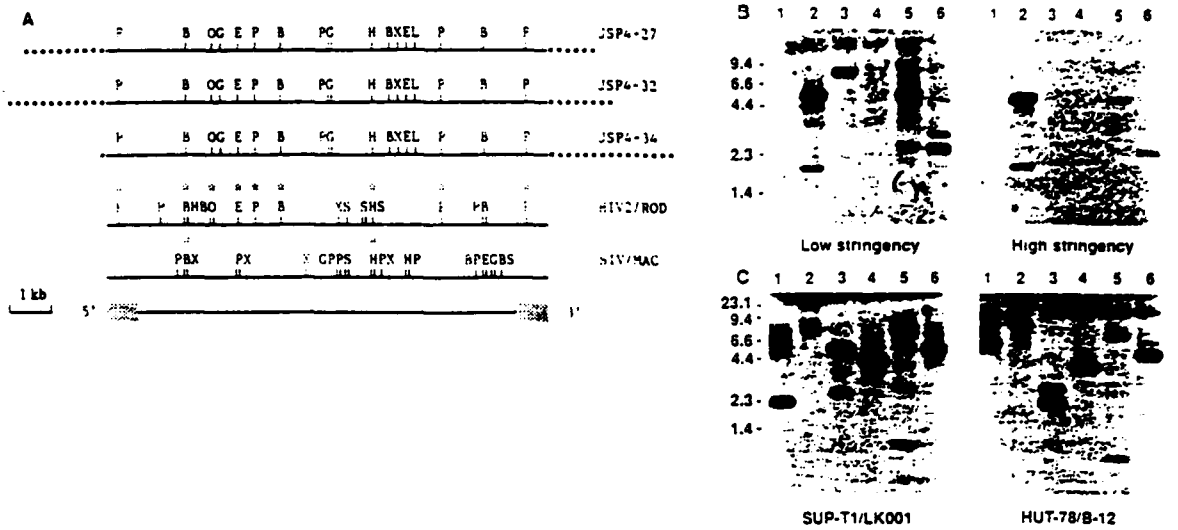


Figure 20. Genetic characterization of HIV-2/ST by differential nucleic acid hybridization, restriction enzyme analysis of HIV-2/ST DNA, and comparison of full-length HIV-2/ST, HIV-2/ROD, and SIV_{MAC} proviral DNA clones. Identical nitrocellulose filters containing agarose gel separated Bam HI restriction digests of virally-infected cellular DNA (10ug each) were prepared (B). HIV-1/IIIb (lanes 1); HIV-2/ROD (lanes 2); SIV_{MAC}-251 (lanes 3); normal human lymphocyte DNA; HIV-2/ST (cell line SupT1/LK001) (lanes 5); and HIV-2/ST (cell line Hut78/B12) (lanes 6). Hybridization was to a ^{32}P-labelled 4.3 kb Pst I-Pst I (*pol*-central region-*env* fragment of HIV-2/ROD probe (Fig. 1C) and filters were washed at low (3 x SSC, 0.2% SDS, 55°C) or high (0.1 x SSC, 0.2% SDS, 65°C) stringency. DNA blot-hybridization restriction cleavage analysis with the same probe as in (B). SupT1/LK001 and Hut 78/B12 are permanent producing HIV-2/ST infected cell lines, the latter derived by single cell cloning of a primary Hut 78 infected line. Restriction enzymes were Xba I (lanes 1), Xho I (lanes 2), Bam HI (lanes 3), Eco RI (lanes 4), Hinc II (lanes 5), and Kpn I (lanes 6). (A) Restriction maps of three full-length recombinant proviral DNA clones of HIV-2/ST (JSP4-27; JSP4-32; JSP4-34) obtained by Sau3AI (Mbo I) partial digestion of Hut 78/B12 DNA and ligation into J1-lambda phage. Restriction enzymes shown are Pst I (P), Bam HI (B), Xho I (O), Bgl II (G), Eco RI (E), Hind III (H), Xba I (X), Sal I (L), and Sst I (S). Asterisks denote sites in HIV-2/ROD or SIV_{MAC} that are also present in the HIV-2/ST clones. Dotted lines denote unique flanking cellular sequences in the HIV-2/ST clones.

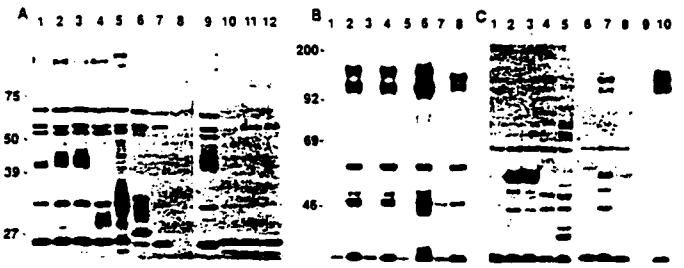


Figure 21. (A) Western immunoblot and (B and C) RIP analysis of HIV-2/ST, HIV-2/ROD, SIV_{MAC}, and HIV-1/IIIb. (A) Immunoblots were performed using a human (West African) antiserum to HIV-2 (lanes 1-8) and a human antiserum to HIV-1 (lanes 9-12). Bound antibody was detected with peroxidase-conjugated goat antibody to human immunoglobulin. Antigen preparations were viral lysates from HIV-2/ST line SupT1/LK001 (lanes 1, 12); single cell derived clones from SupT1/LK001 designated ST.17 (lane 2), ST.9 (lane 3), ST.24 (lanes 4, 11); HIV-2/ROD (lanes 5, 10); SIV_{MAC} (lanes 6); HIV-1/IIIb (lanes 7, 9); uninfected control cells (lane 8). (B) RIP of HIV-2/ST *env* proteins from cloned cell lines of infected SupT1 cells shown in (A). Cell lysates from [^{35}S]cysteine and [^{35}S]methionine labeled clones were prepared and immunoprecipitated with either normal human serum (lanes 1, 3, 5, 7) or serum from a West African patient with HIV-2 infection (lanes 2, 4, 6, 8). Shown are clones designated ST.9 (lanes 1 and 2), ST.17 (lanes 3 and 4), ST.23 (lanes 5 and 6) that have an *env* precursor molecule of approximately 180 kD and ST.24 (lanes 7 and 8) which has an *env* precursor molecule of 170 kD. The smaller sized *env* precursor (gp170) of ST.24 corresponds to the truncated transmembrane protein (gp30) of the same virus (Fig. 1A). (C) Co-precipitation of HIV-2/ST envelope glycoproteins with CD4. Cell lysates from [^{35}S]cysteine and [^{35}S]methionine-labeled uninfected SupT1 cells (lanes 1-5) and HIV-2/ST infected SupT1/LK001 cells (lanes 6-10) were prepared and immunoprecipitated with OKT3 (lanes 1 and 6), OKT4 (lanes 2 and 7), OKT4A (lanes 3 and 8), normal human serum (lanes 4 and 9), or human serum from a patient with HIV-2 infection (lanes 5 and 10). Both OKT4 and OKT4A immunoprecipitate the 56kD CD4 molecule from the uninfected cells. However, in the infected cells OKT4 (but not OKT4A) immunoprecipitates CD4 as well as three molecular species of high molecular weight corresponding to viral *env* gene products that are also immunoprecipitated by immune human serum (lane 10).

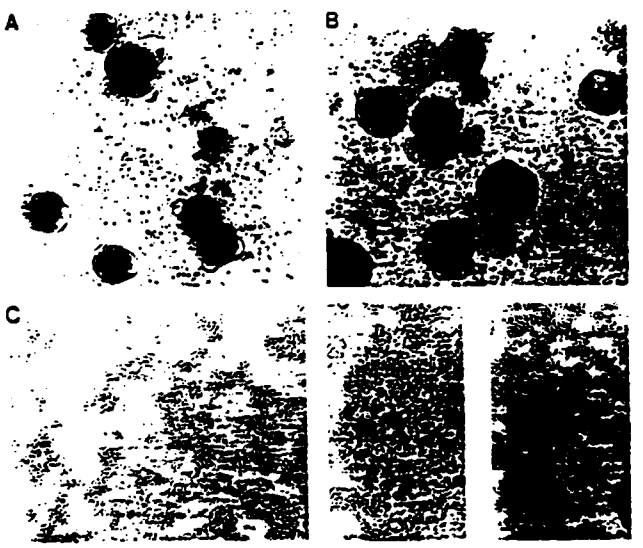


Figure 22. Syncytium formation by HIV-2/ST, HIV-2/ROD, and HIV-1/IIIb infected cells and CD4⁺ SupT1 indicator cells. Virally-infected SupT1 cells were mixed 1:5 with uninfected SupT1 cells, incubated 24 hours, and photographed. Results were identical using Hela-T4, H9, and Hut 78 as uninfected indicator cells and when using HIV-2/ST infected H9 and Hut 78 as *env* expressing effector cells. The HIV-2/ST cell line (ST/LK001) and single cell derived clones (ST.9 and ST.24) are described in the legend to Fig. 21.

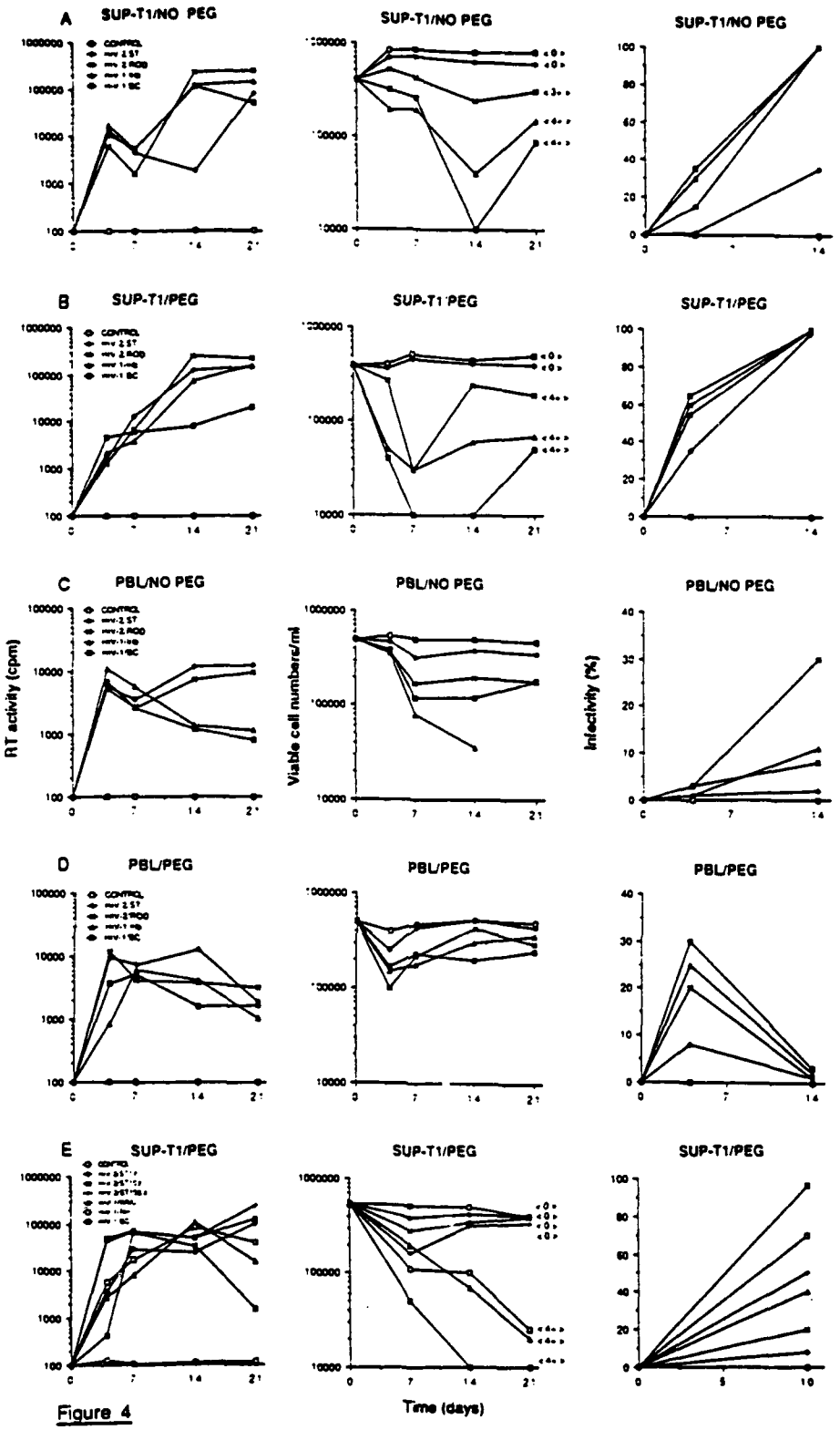


Figure 4

Figure 23. Comparison of cytopathic/cell killing properties of HIV-2/ST with HIV-2/ROD and with HIV-1 isolates IIIb, BC, WMJ, RH. Each panel summarizes a 21 day experiment in which equivalent amounts of cell-free virus (250,000 c.p.m. RT activity) were applied to cultures of uninfected SupT1 cells (A, B) or PHA-stimulated peripheral blood lymphocytes (Panels C, D). In Panel E, 7,500 c.p.m. RT activity of all viruses were used except for HIV-2/ST 10X (75,000 c.p.m.) and HIV-2/ST 100X (750,000 c.p.m.). Shown are supernatant reverse transcriptase activities; percentages of cells infected with virus as determined by indirect immunofluorescence; and cell killing assessed by viable cell counts with trypan blue exclusion, hemocytometry, and automated cell counting. Identical culture splits were done weekly to keep the concentration of control cells between 0.5 x 10^6 and 1.0 x 10^6/ml and cell killing was expressed logarithmically as decreases in viable cell numbers. Syncytia were scored from 0 (none) to 4+ (extensive) and are shown in brackets (< >). Mock infected culture supernatants lacking virus served as controls. Ten percent (w/v) final concentration of polyethylene glycol (PEG) was used to concentrate virus for experiments shown in Panels B, D, and E.

Isolate	Cells expressing viral mRNA (%)	Viral mRNA per cell
	Day 1 after infection	
HIV-1/IIIb	3	++++
HIV-2/ROD	8	++++
HIV-2/ST	0.01	++++
Control	0	-
	Day 3 after infection	
HIV-1/IIIb	35	++++
HIV-2/ROD	45	++++
HIV-2/ST	0.01	++++
Control	0	-
	Day 7 after infection	
HIV-1/IIIb	100	++++
HIV-2/ROD	100	++++
HIV-2/ST	0.5	++++
Control	0	-
	Day 10 after infection	
HIV-1/IIIb	100	++++
HIV-2/ROD	100	++++
HIV-2/ST	10	++++
Control	0	-
	Day 14 after infection	
HIV-1/IIIb	ND	ND
HIV-2/ROD	ND	ND
HIV-2/ST	50	++++
Control	0	-

Figure 24. Time course of viral infection by HIV-1_{IIIb}, HIV-2_{ROD}, and HIV-2_{ST} assessed by in situ hybridization. The relative amount of viral mRNA production per cell was scored qualitatively from silver grain densities: -, absent, to ++++, too numerous to count. Although there was a marked delay in the initial development and spread of productive viral infection by HIV-2_{ST}, the few cells that became productively infected at early time points expressed equal amounts of viral RNA on a per cell basis compared with prototype HIV-1 and HIV-2

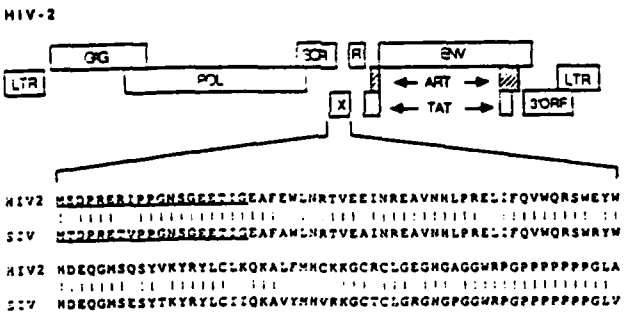


Figure 25. Alignment of the deduced X amino acid sequence of HIV-2 and SIV_{MAC}. Shown are eight open reading frames known to encode HIV proteins as well as the novel X open reading frame. A comparison of the deduced X amino acid sequence is shown below for HIV-2 (isolate ROD) and SIV_{MAC} (isolate PK82). Vertical bars indicate sequence identity. The first 19 residues (underlined in the HIV-2 and SIV_{MAC} X sequence) represent the oligopeptides which were synthesized to raise rabbit anti-X immune sera.

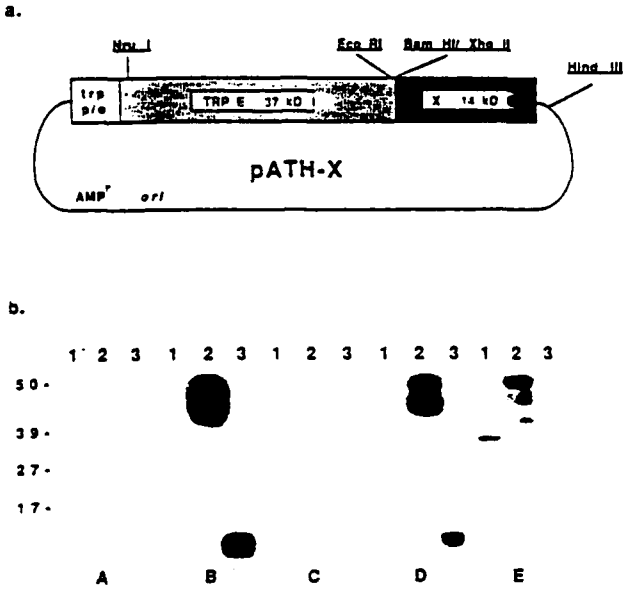


Figure 26. Prokaryotic expression of the SIV_{MAC} X protein. (A) Construction of pATH-X. The prokaryotic expression vector, pATH, was used to express the entire SIV_{MAC} X open reading frame minus the first two amino acid residues. A 422 base pair XhoII/HindIII fragment of SIV_{MAC} clone PKE102 was inserted into pATH at the BamHI and HindIII sites as depicted. This placed the SIV_{MAC} orf inframe and downstream of the bacterial Trp-E gene to allow its expression as a 50kD fusion protein. A unique NruI and EcoRI site were subsequently used to remove 95% of the Trp-E coding region, leaving only 16 TrpE derived amino acids on the 5' end of the expressed SIV_{MAC} X gene. This construct yielded a smaller fusion protein of 15kD. The position of the trp promoter/operon within the construct is indicated. (B) Western blot analysis of bacterially-expressed SIV_{MAC} X proteins. *E. coli* cell lysates transformed with non-recombinant vector pATH (lanes 1), pATH-X (lanes 2) and the TrpE deleted pATH-X construct (lanes 3) were tested by immunoblot analysis for reactivity to rabbit sera immunized with SIV_{MAC} and HIV-2 X peptides (panels B and D, respectively) as well as to preimmune rabbit sera (panels A and C). A 50kD TrpE/X and a 15kD delta TrpE/X fusion protein were recognized only by X peptide immune sera. (A smaller band of approximately 40kD represents a prematurely truncated version of the 50kD TrpE/X protein.) An antiserum raised against the bacterial TrpE gene product (panel E) was used as control to confirm the size of the 50kD TrpE/X fusion protein (lane 2) and to identify the non-recombinant 37kD TrpE protein (lane 1). This antiserum failed to detect the TrpE deleted 15kD X protein (lane 3). Antigen antibody complexes were detected using ^{125}I labelled protein A.

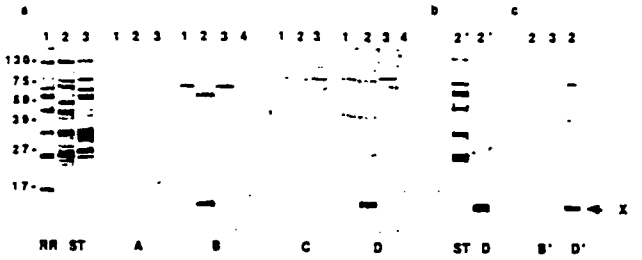


Figure 27. Identification of the HIV-2 and SIV_{MAC} X protein. (A) Immunoblot of viral preparations probed with X peptide immune and preimmune sera. HIV-1 (isolate IIIb, lanes 1), HIV-2 (isolate ST, lanes 2), and SIV_{MAC} viral lysates (isolate PK82, lanes 3), and lysates of uninfected cell culture supernatants (lanes 4) were separated on a 12.5% polyacrylamide-sodium dodecyl sulfate gel, electrophoretically transferred to nitrocellulose, and reacted to the following sera diluted 1:100 in PBS containing 5% nonfat dry milk and 0.1% Tween 20: human anti-HIV-1 positive serum (RR); human anti-HIV-2 positive serum (ST); rabbit SIV X peptide preimmune (A) and immune serum (B); rabbit HIV-2 X peptide pre-immune (C) and immune serum (D). Bound antibody was detected by reaction with horseradish peroxidase conjugated anti-rabbit and anti-human immunoglobulins using diamianobenzidine as a substrate. (B) Immunoblots of sucrose banded, purified HIV-2 virions. HIV-2 viral supernatants (isolate ST) were clarified by low-speed centrifugation, virus was pelleted through a 20% sucrose cushion and then banded in a 20% to 60% continuous sucrose gradient. Immunoblots were prepared and identical strips containing purified HIV-2 virions as antigen (2') were reacted with an anti-HIV-2 positive human serum (ST) and the rabbit HIV-2 X peptide immune sera (D). (C) Competitive adsorption assay using bacterially expressed SIV_{MAC} X protein to adsorb anti-X antibodies. Bacterial lysates of pATH-X transformed *E. coli* immobilized on nitrocellulose were incubated with SIV_{MAC} and HIV-2 X peptide immune sera (panel B' and D', respectively). Preadsorbed sera were subsequently tested by Western blot analysis for reactivity with preparations of HIV-2 (isolate ST, lanes 2) and SIV_{MAC} (isolate PK82, lane 3).

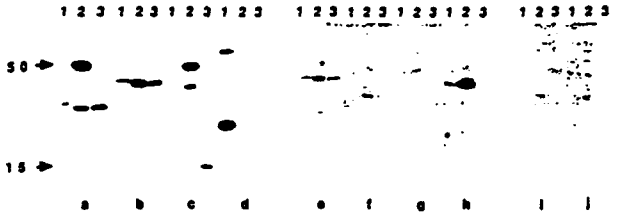


Figure 28. Analysis of human sera for reactivity with the SIV_{MAC} X proteins expressed in *E. coli*. Cell lysates transformed with pATH vector (lanes 1), pATH-X (lanes 2) and the TrpE deleted pATH-X construct (lanes 3) were separated on a 12.5% acrylamide sodium dodecyl sulfate gel, transferred to nitrocellulose and analyzed by immunoblotting for reactivity to four HIV-2 antibody positive human sera (a-d), four HIV-1 positive human sera (e-h) and 2 normal control sera (i, j). Sera were diluted 1:50 in PBS containing 5% non-fat dry milk and 0.1% Tween 20. Arrows denote the position of the 50kD TrpE/X and the 15kD delta TrpE/X fusion proteins recognized by two of four HIV-2 positive human sera (a and c). Bound antibody was detected by ^{125}I labelled protein A.

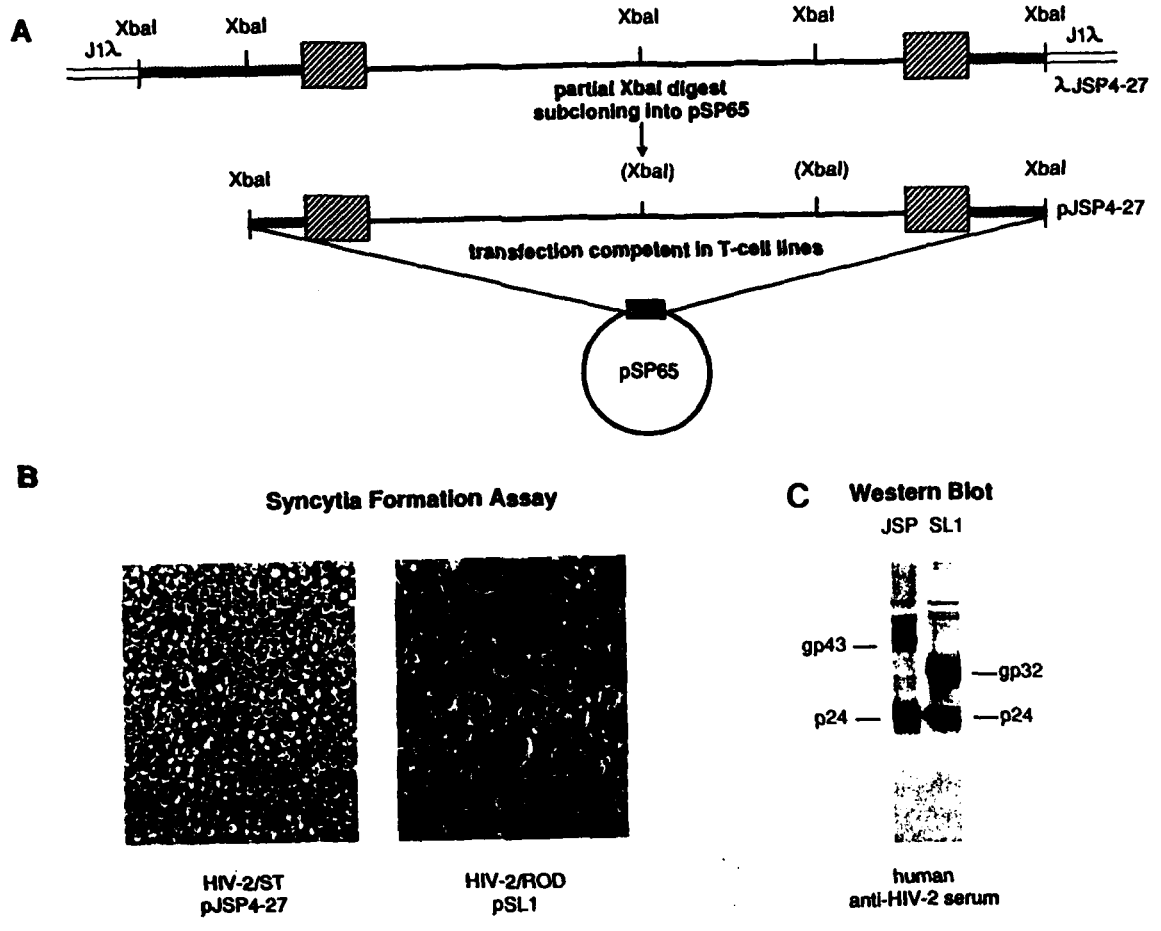


Figure 29. (A) Construction of a replication-competent HIV-2/ST plasmid clone. λJSP4-27 was partially cleaved with *XbaI* to remove bacteriophage arms (double lines) and flanking cellular sequences (hatched lines), and the resulting 14-kilobase-pair provirus-containing fragment was subsequently subcloned into pSP65. (B and C) Biological comparison of HIV-2/ST- and HIV-2/ROD-derived, genetically pure viral strains. CEMx174 cultures, productively infected (>90%) with JSP4-27 (HIV-2/ST) and SL1 (HIV-2/ROD), respectively, were examined in syncytium formation assays and by Western blot analysis. (B) No syncytium formation was observed upon cocultivation of JSP4-27-producing CEMx174 cells with uninfected CEM cells, whereas numerous and large syncytia were generated upon cocultivation of the same uninfected CEM cells with SL1-infected CEMx174 (identical results were obtained with H9, SupT1, and CEMx174 cells). Syncytium formation was monitored 18 h after cocultivation. (C) Western blot analysis of cell-free virions derived from these same transfection-derived cultures demonstrated differences in the sizes of the JSP and SL1 transmembrane envelope glycoproteins.

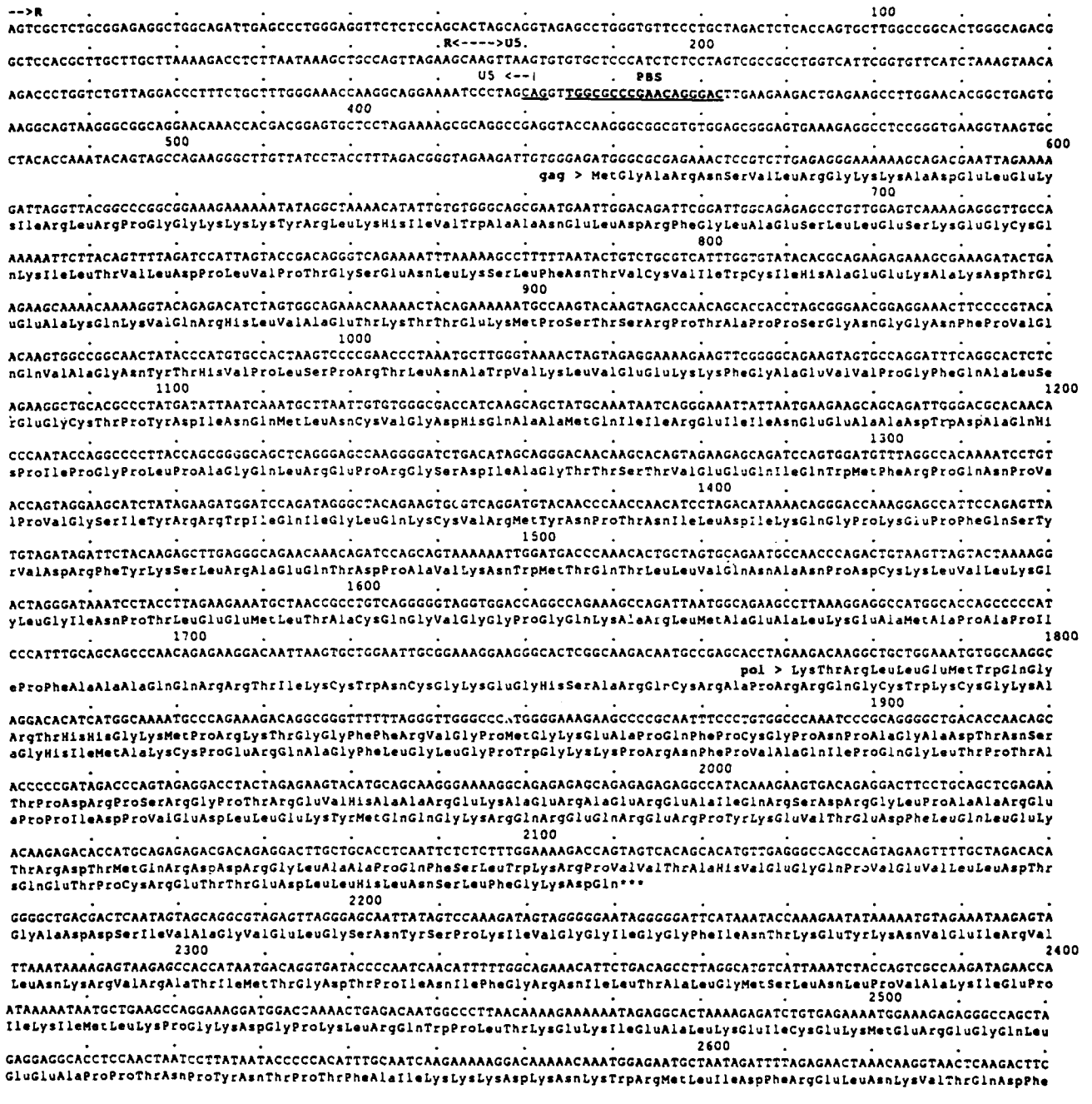


Figure 30. Complete nucleotide sequence of the HIV-2/ST proviral genome. Shown are 9,762 bp of nucleotide sequence and the deduced amino acid sequences of the corresponding viral proteins. The sequence starts at the 5' cap site and ends with the 3' polyadenylation site of the viral RNA. The primer-binding site (PBS; complementary to the Lys-tRNA), polypurine tract (PPT), and short inverted repeats that flank the LTRs are underlined. Core enhancer sequences (E), Sp1-binding sites (Sp1), the TATA box (TATAA), and the polyadenylation signal (AATAAA) are shown. The U3-R and R-U5 boundaries, as well as the splice donor (SD) and splice acceptor (SA) sites, have been determined in analogy with HIV-2/ROD. The *vpr* open reading frame contains a premature in-frame TAA stop codon at position 5777 ("***"). Sequence analysis was performed by the chemical degradation method of Maxam and Gilbert as well as by the dideoxynucleotide-chain termination method of Sanger et al. The nucleotide sequence of JSP4-27 has been submitted to the AIDS Sequence Data Base, Los Alamos National Laboratories, as well as to the GenBank and EMBL Laboratories.

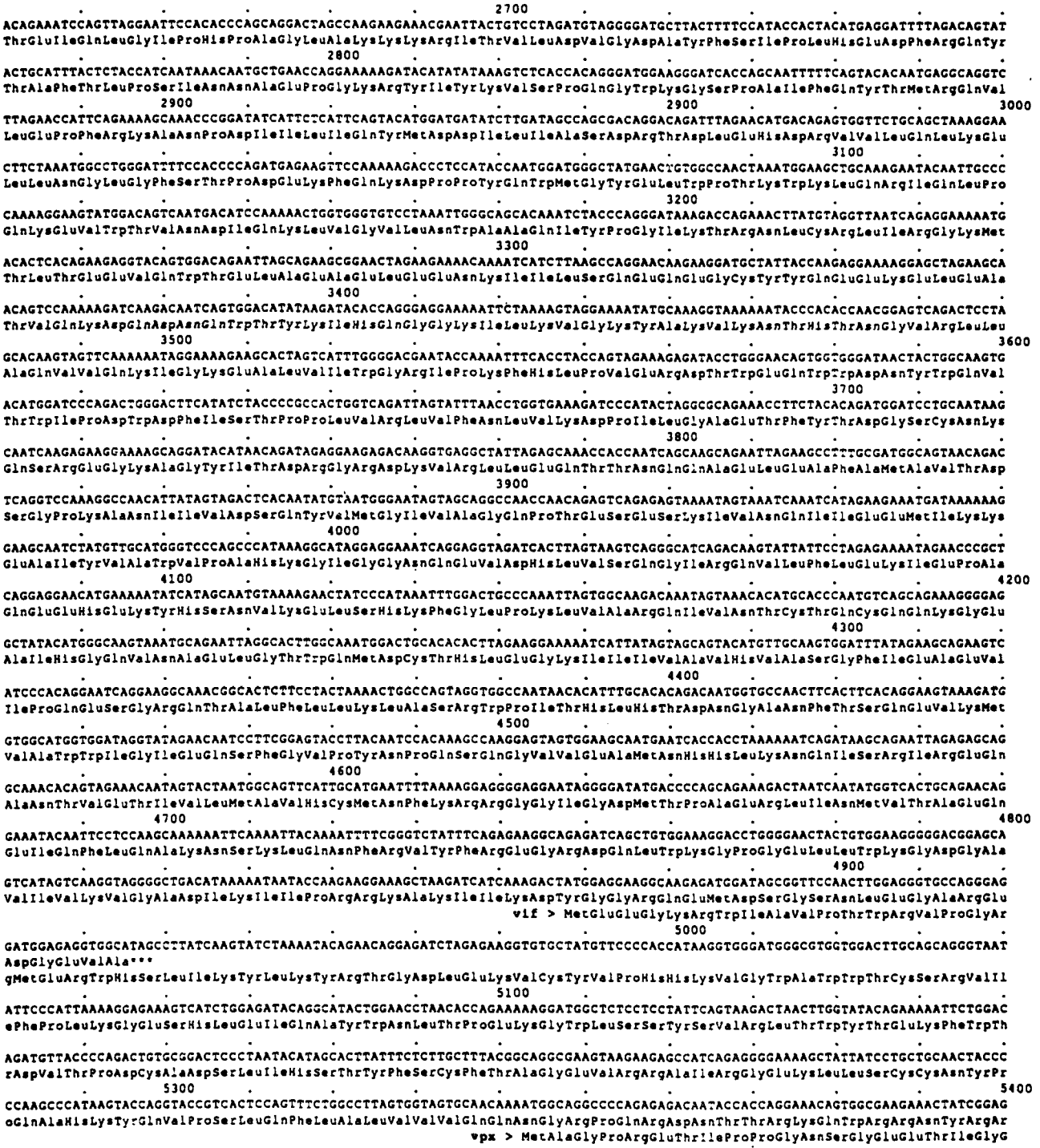


Figure 30 (cont.)

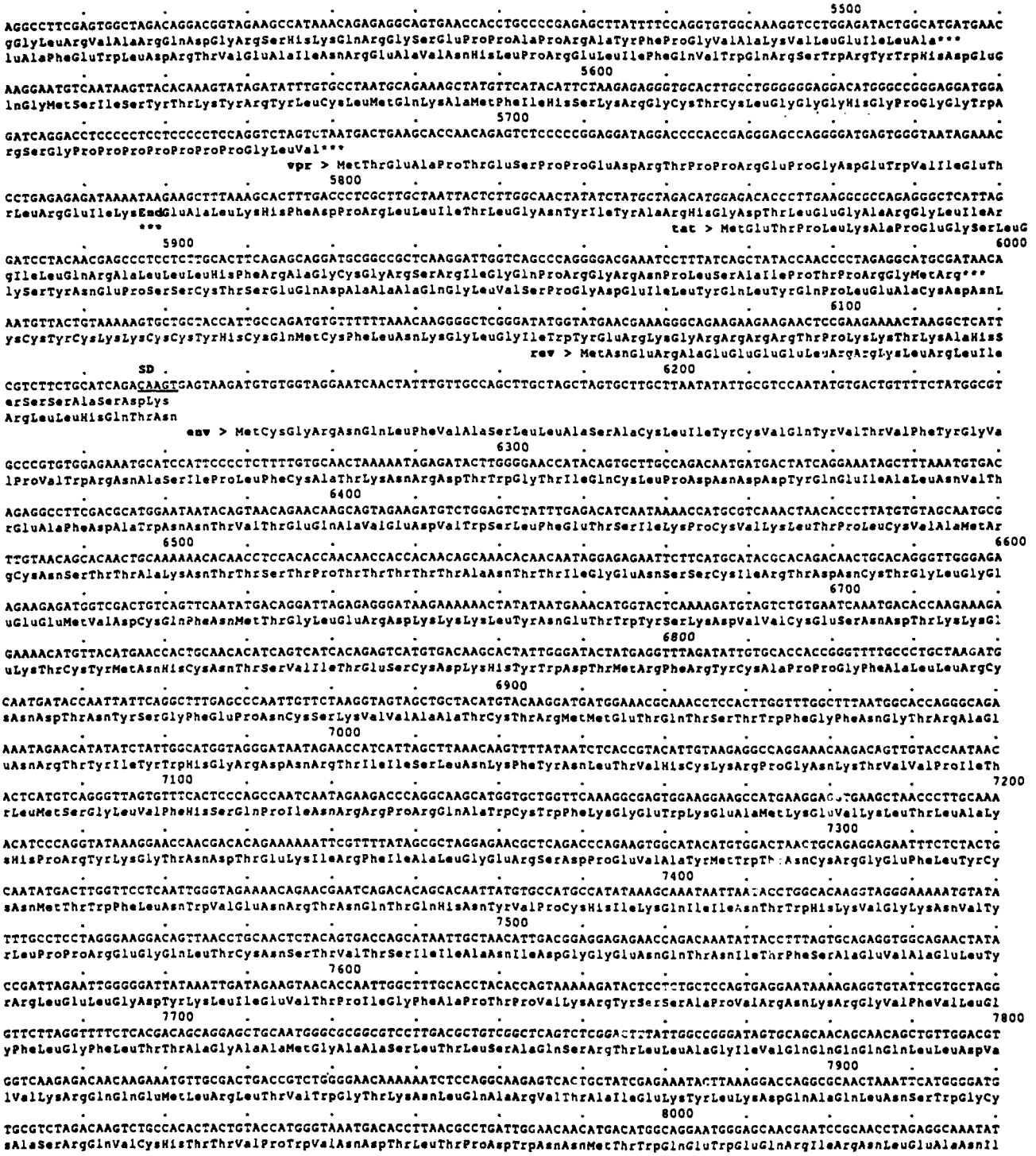


Figure 30 (cont.)

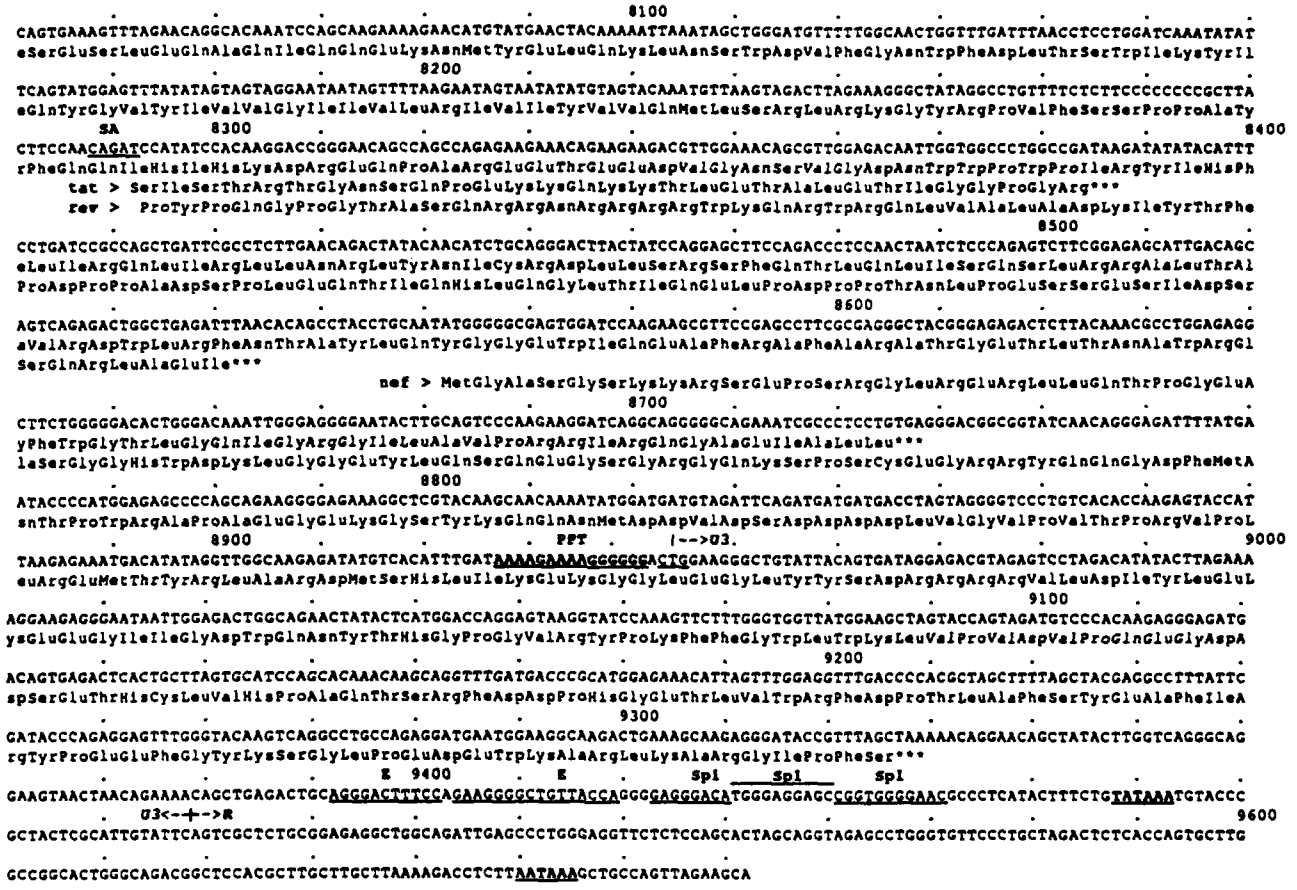


Figure 30 (cont.)

	Envelope Amino Acid Sequence Divergence						
	HIV-2/ST	HIV-2/ROD	HIV-2/ISY	HIV-2/NIH$_z$	HIV-2/GH	SIV/MAC$_{142}$	SIV/SM
HIV-2/ST		19%	17%	18%	16%	28%	28%
HIV-2/ROD	11%		20%	19%	18%	28%	28%
HIV-2/ISY	10%	11%		20%	19%	30%	29%
HIV-2/NIH$_z$	12%	12%	13%		19%	28%	28%
HIV-2/GH	11%	12%	12%	15%		29%	29%
SIV/MAC$_{142}$	23%	23%	24%	26%	24%		19%
SIV/SM	23%	22%	23%	23%	23%	15%	

Total Nucleotide Sequence Divergence

Figure 31. Nucleotide and amino acid sequence divergence among HIV-2 and SIV strains. The percent nucleotide sequence divergence between HIV-2/ST (JSP4-27), HIV-2/ROD, HIV-2/ISY, HIV-2/GH, SIV$_{MAC142}$, and SIV$_{SM}$ is shown along with the percent amino acid sequence divergence of their envelope glycoproteins. Sequences were aligned pairwise, using the Microgenie computer software (Beckman).

Open reading frame	% Homology					
	ST/ROD		ST/ISY		ROD/ISY	
	Nucleotide	Amino acid	Nucleotide	Amino acid	Nucleotide	Amino acid
LTR	91.5		91.4		91.8	
gag	91.1	92.0	90.7	89.1	90.3	89.5
pol	91.2	91.4	91.2	91.4	91.2	91.7
vif	91.5	88.4	92.4	91.2	92.2	85.6
vpx	89.3	85.7	92.6	90.2	88.4	87.5
vpr	89.9[a]	80.0[a]	89.8[a]	84.8[a]	93.3	89.5
tat	86.7	75.4	88.7	78.5	90.5	80.8
rev	82.0	85.0	ND[b]	ND[b]	ND[b]	ND[b]
env	85.5	81.4	86.0	83.0	84.8	80.4
nef	86.1	78.9	85.8	77.7	87.0	78.5
Overall % homology	89.5		89.9		89.3	

Figure 32. Sequence homologies among virus-specific genes of three HIV-2 proviruses.

[a]An in-frame stop codon is present in the HIV/S/ST *vpr* open reading frame.

[b]Meaningful comparison of the HIV-2/ISY *rev* gene with the corresponding *rev* genes of HIV-2/ST (JSP4-27) and HIV-2/ROD was not possible because of considerable length differences between their sequences. ND, Not done.

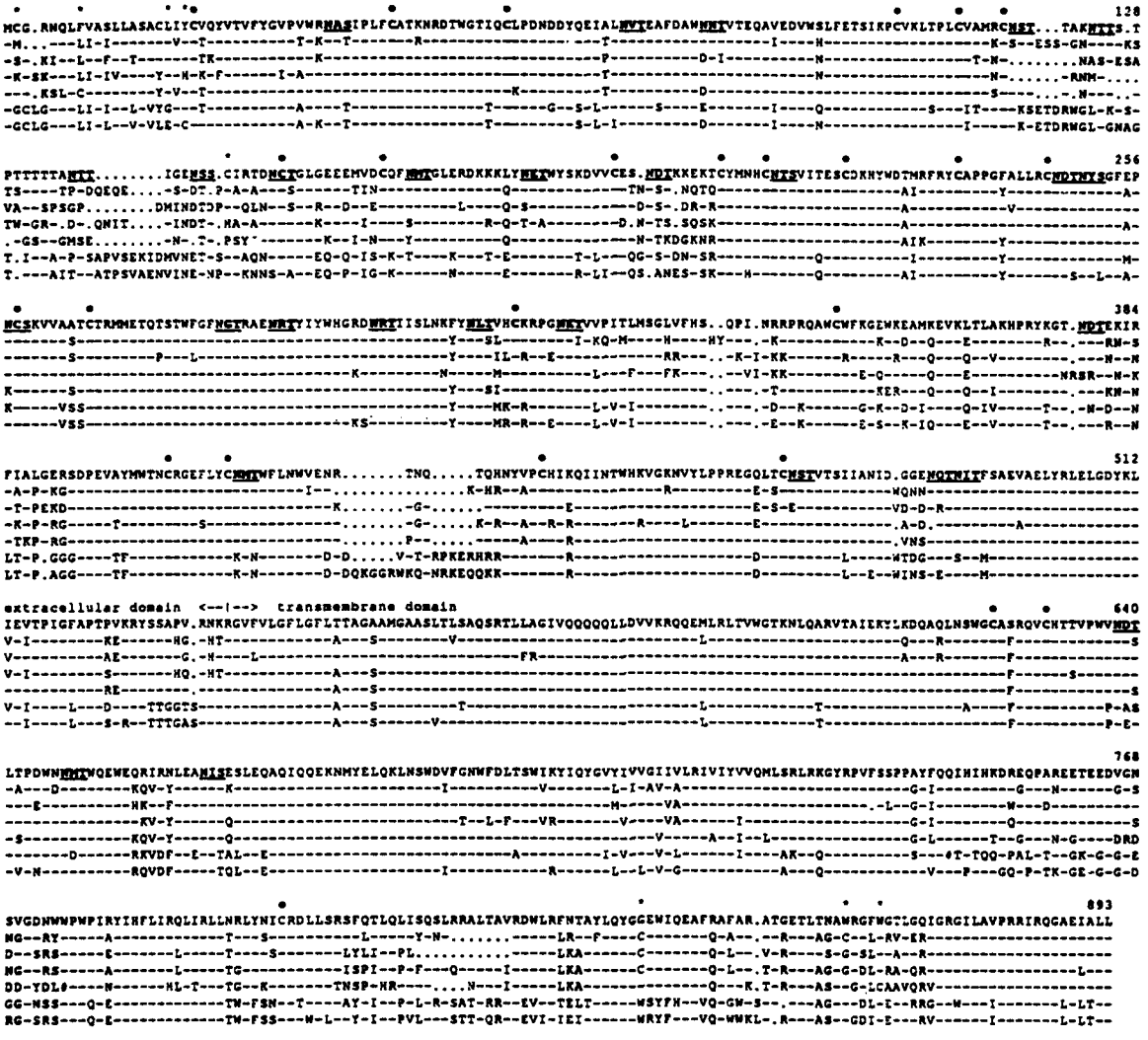


Figure 33: Alignment of the deduced amino acid sequences of the envelope open reading frames of HIV-2/ST (JSP4-27), HIV-2/ROD, HIV-2/ISY, HIV-2/NIHz, HIV-2/GH, SIVMAC and SIVSM. The position of the presumed primary envelope precursor cleavage site, as identified for HIV-2/ROD, is shown. Symbols: ●, cysteine residues conserved among all seven envelope sequences; *, nonconserved cysteine residues; #, in-frame stop codons present in the sequences of SIVMAC and HIV-2/GH; •, gaps introduced for optimal sequence alignment. Potential N-linked glycosylation sites (NXS/T) in the HIV-2/ST envelope sequence are underlined.

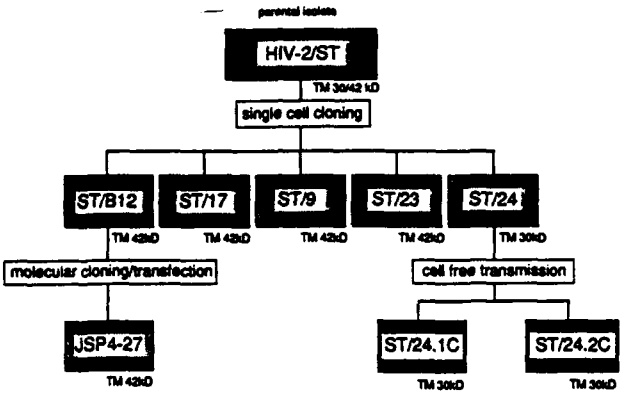


Figure 34. Diagram of the generation of fusogenic and cytopathic variants of HIV-2/ST showing the origins of biologically distinct from the original HIV-2/ST isolate by limiting dilution cloning. All cell line, and the parental bulk culture produce nonfusogenic and noncytopathic progeny virus (▒). Two fusogenic and cytopathic variants ST/24.1C and ST/24.2C were generated by repeated cell free passage of ST/24 supernatants to uninfected SupT1 cells (■). The size of the envelope transmembrane glycoprotein (TM) for virions derived from each culture is shown. kD, Kilodaltons.

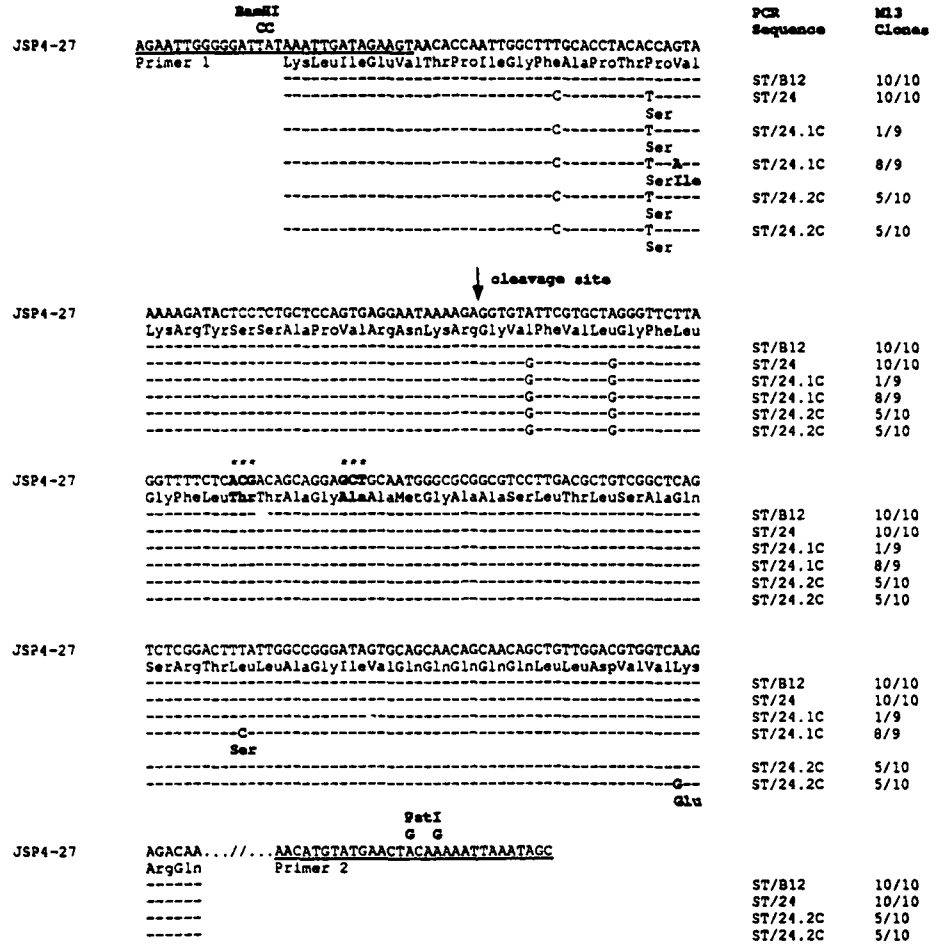


Figure 35. Nucleotide sequence variation in the fusion domain of cytopathic and noncytopathic HIV-2/ST strains. A stretch of 230 bp of PCR-amplified envelope sequence is compared between JSP4-27 and four other HIV-2/ST strains. The boundaries of the amplified fragments are indicated. The sequences of both primers used to amplify the genomic DNA of ST/B12, ST/24, ST/24.1C and ST/24.2C are underlined, and the base pair changes that were introduced to generate *Bam*HI and *Pst*I cloning sites are indicated. Nucleotide substitutions are shown with respect to the JSP4-27 sequence, and amino acid sequence changes unique to the cytopathic and fusogenic ST/24 strains are in boldface. Asterisks mark the threonine and alanine substitutions previously identified to distinguish JSP4-27 from most other cytopathic HIV-2 strains. The number of M13 clones analyzed per HIV-2/ST strain is listed, with frequencies referring to the proportion of clones that have identical sequences.

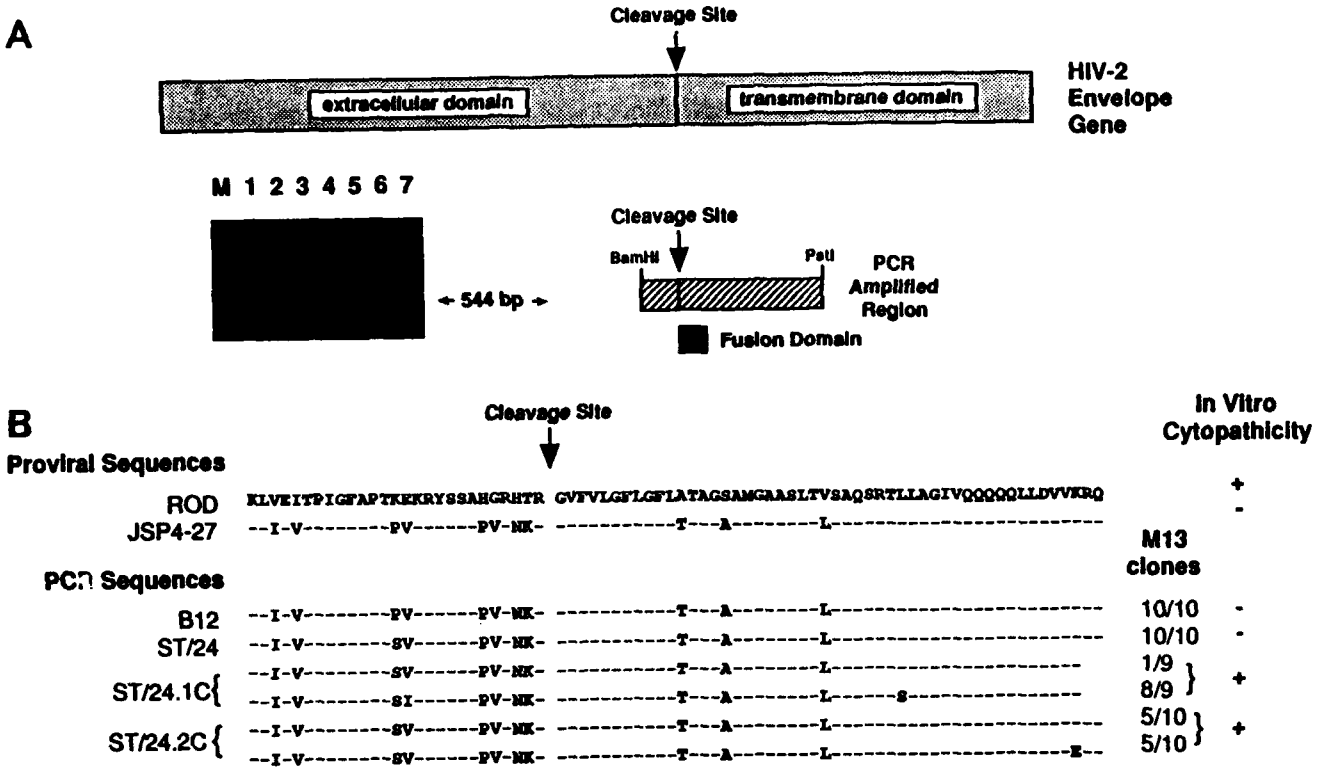


Figure 36. Amino acid sequence variation in the fusion domain of cytopathic and noncytopathic HIV-2/ST strains. (A) Relative locations of the PCR-amplified envelope fragments in the context of the entire HIV-2 envelope open reading frame. Amplification products: Lanes: 1 and 7, uninfected peripheral blood lymphocyte DNA (negative control); 2, ST/B12; 3, ST/24; 4, ST/24.1C; 5, ST/24.2C; 6, SupT1/LK001 (HIV-2/ST-infected positive control cell line). (B) Alignment of the deduced amino acid sequences of the amplified fusion regions to the corresponding sequences of HIV-2/ROD and JSP4-27. the number of M13 clones analyzed per HIV-2/ST strain is listed, with frequencies referring to the proportion of clones that have identical sequences.

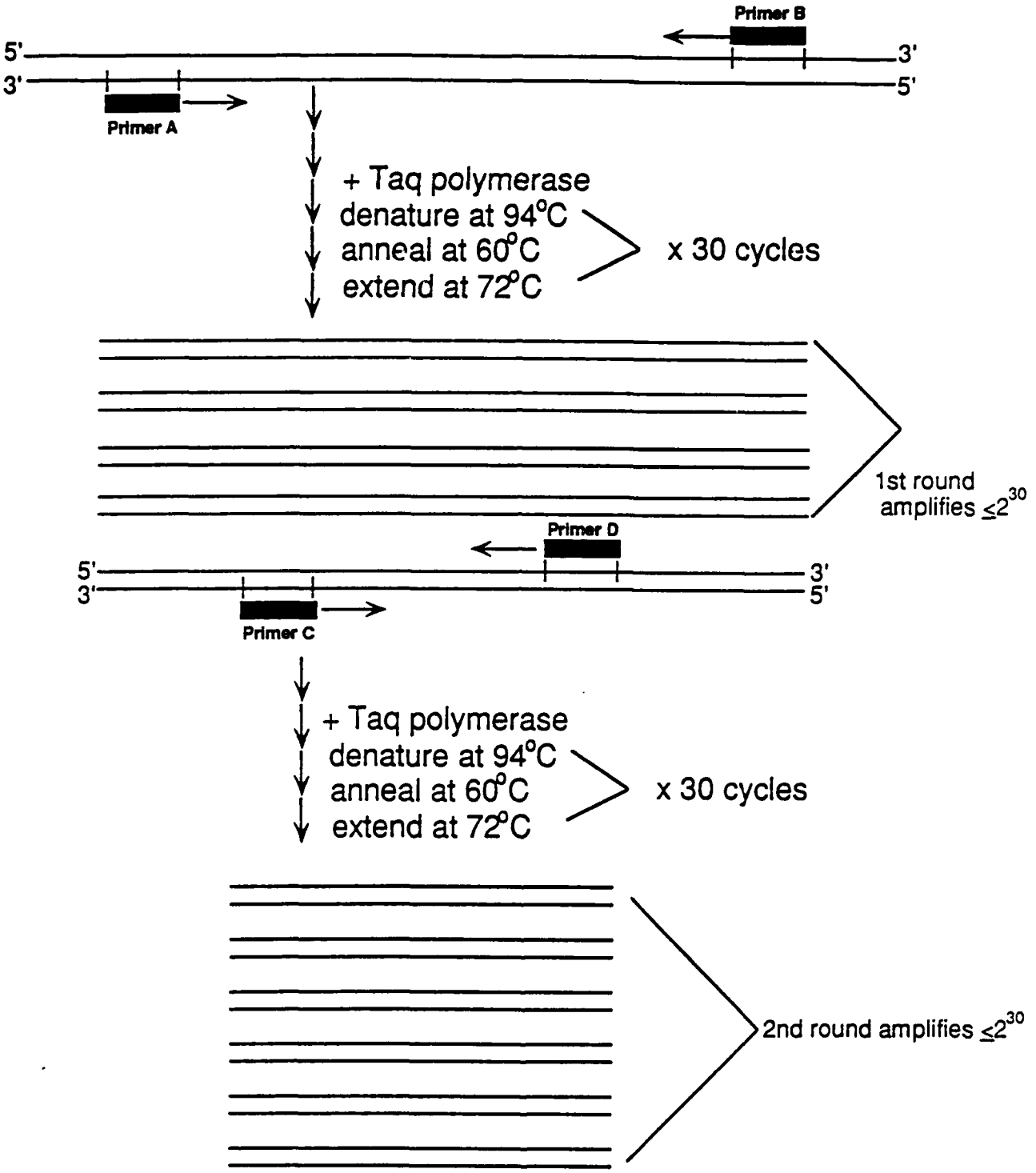


Figure 37. "Nested" polymerase chain reaction primer amplification as an approach to increase sensitivity of retroviral genome detection.

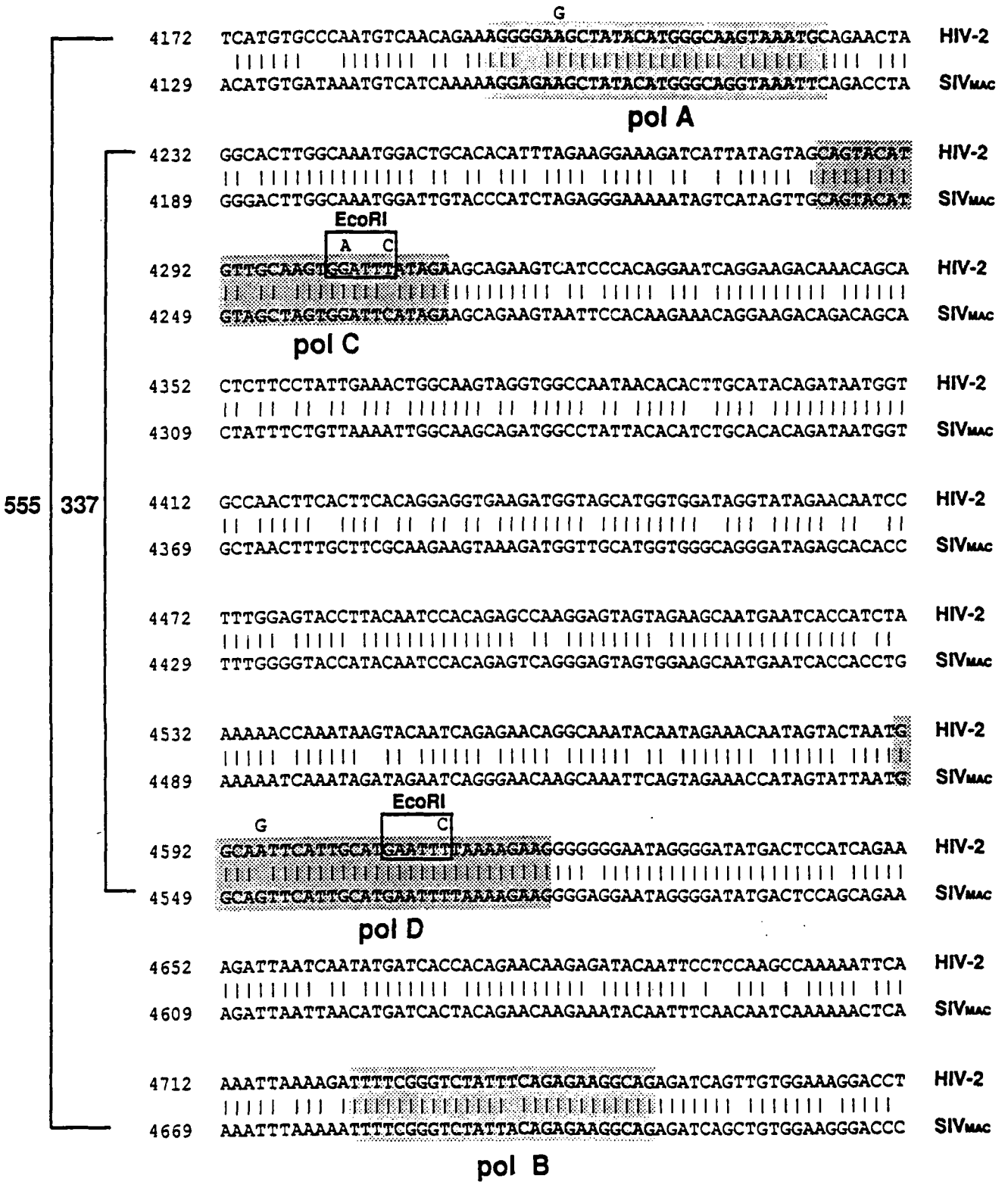


Figure 38. Nested HIV-2/SIV$_{MAC}$ pol primers with cloning enzyme sites. Outer (pol A and D) and inner (pol B and C) oligonucleotide primer sequences are shown in shaded areas and are derived from best match sequence of HIV-2 and SIV$_{MAC}$. EcoRI sites were introduced within the inner primer sequences to facilitate cloning of the amplified product. The size of the outer and inner amplification products are 555bp and 337bp, respectively.

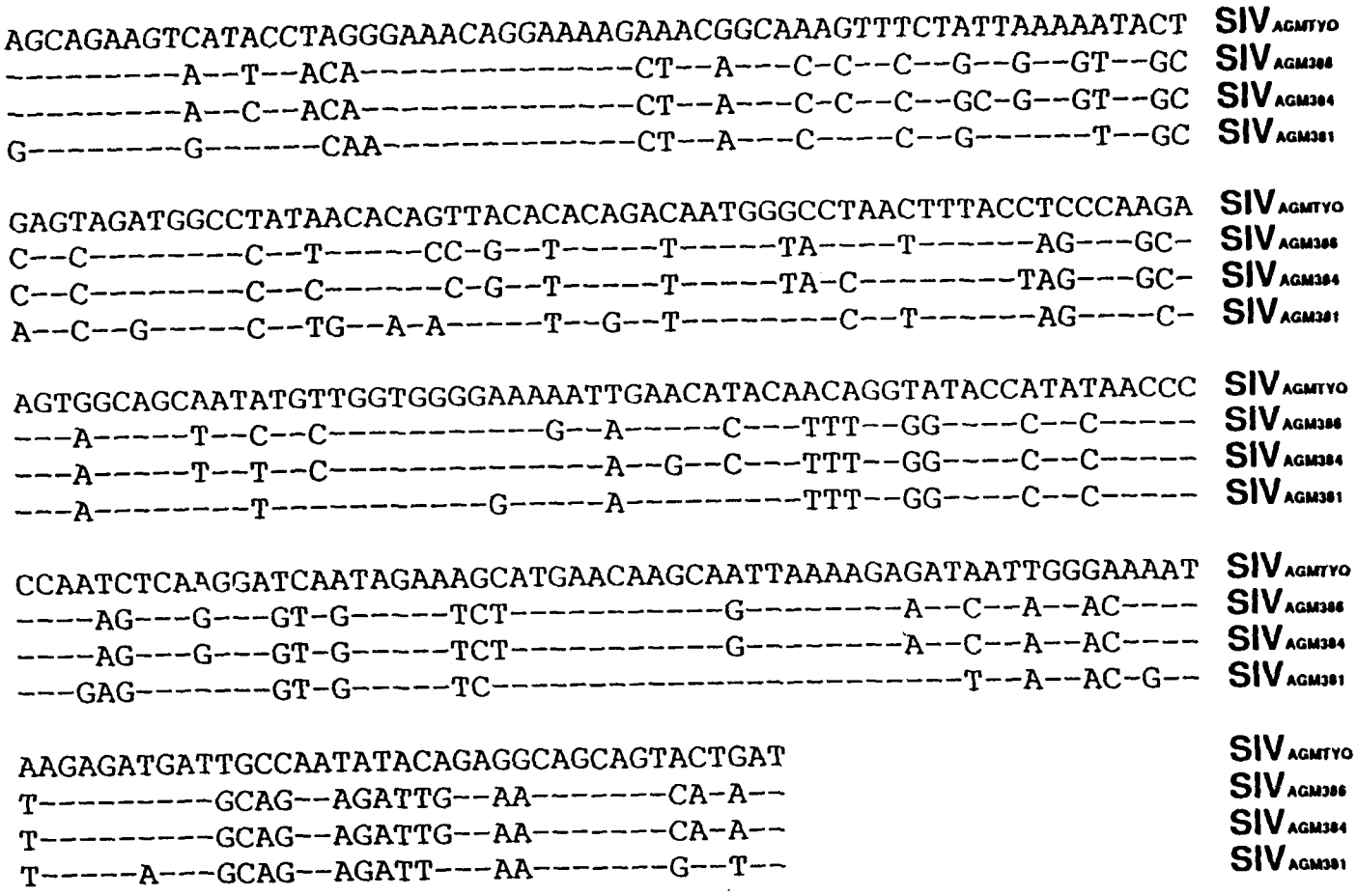


Figure 39. Nucleotide sequence comparison of pol gene segments of East (SIV_{AGMTYO}) and West ($SIV_{AGM386, 384, 381}$) African green monkey viruses. Nucleotide sequences were determined by the dideoxy chain termination method from three West African SIV_{AGM} virus strains (sabeus subspecies) and are shown compared to an East African strain (SIV_{AGM}-Tyo).

SIV_{AGM} SAB-1	SIV_{AGM} SAB-2	SIV_{AGM} SAB-3	SIV_{AGM} SAB-4	SIV_{AGM} TYO	SIV_{AGM} 155	SIV_{AGM} VER-2	SIV_{AGM} GRI-1	SIV/MND	SIV/MAC	SIV/SM	HIV-2	HIV-1	
--	96.7	96.9	93.5	80.4	78.3	81.5	80.4	69.6	69.6	69.6	68.5	70.7	SIV_{AGM}/SAB-1
	---	96.7	94.6	80.4	79.3	81.5	80.4	69.6	69.6	69.6	68.5	69.6	SIV_{AGM}/SAB-2
		---	93.5	79.3	79.3	80.4	81.5	69.6	70.7	70.7	69.6	70.7	SIV_{AGM}/SAB-3
			---	79.3	78.3	80.4	79.3	72.8	70.7	70.7	69.6	69.6	SIV_{AGM}/SAB-4
				---	92.4	93.5	80.4	67.4	62.0	62.0	60.9	62.0	SIV_{AGM}/TYO
					---	94.6	79.3	67.4	65.2	65.2	62.0	62.0	SIV_{AGM}/155
						---	81.5	69.6	67.4	66.3	63.0	64.1	SIV_{AGM}/VER-2
							---	68.5	63.0	63.0	62.0	69.6	SIV_{AGM}/GRI-1
								---	64.1	64.1	62.0	64.5	SIV/MND
									---	95.7	90.2	63.0	SIV/MAC
										---	92.4	65.2	SIV/SM
											---	62.0	HIV-2
												---	HIV-1

Figure 40. Percent amino acid sequence identity in a 279bp PCR amplified pol region of primate lentiviruses. The sequences of the four West African SIV_{AGM} viruses are indicated by $SIV_{AGMsab1-4}$. The four $SIV_{AGM-sab}$ viruses differ from each other by 1-7% but from East African viruses by 19-22% and from isolates of HIV-1 and HIV-2 by 29-31%.

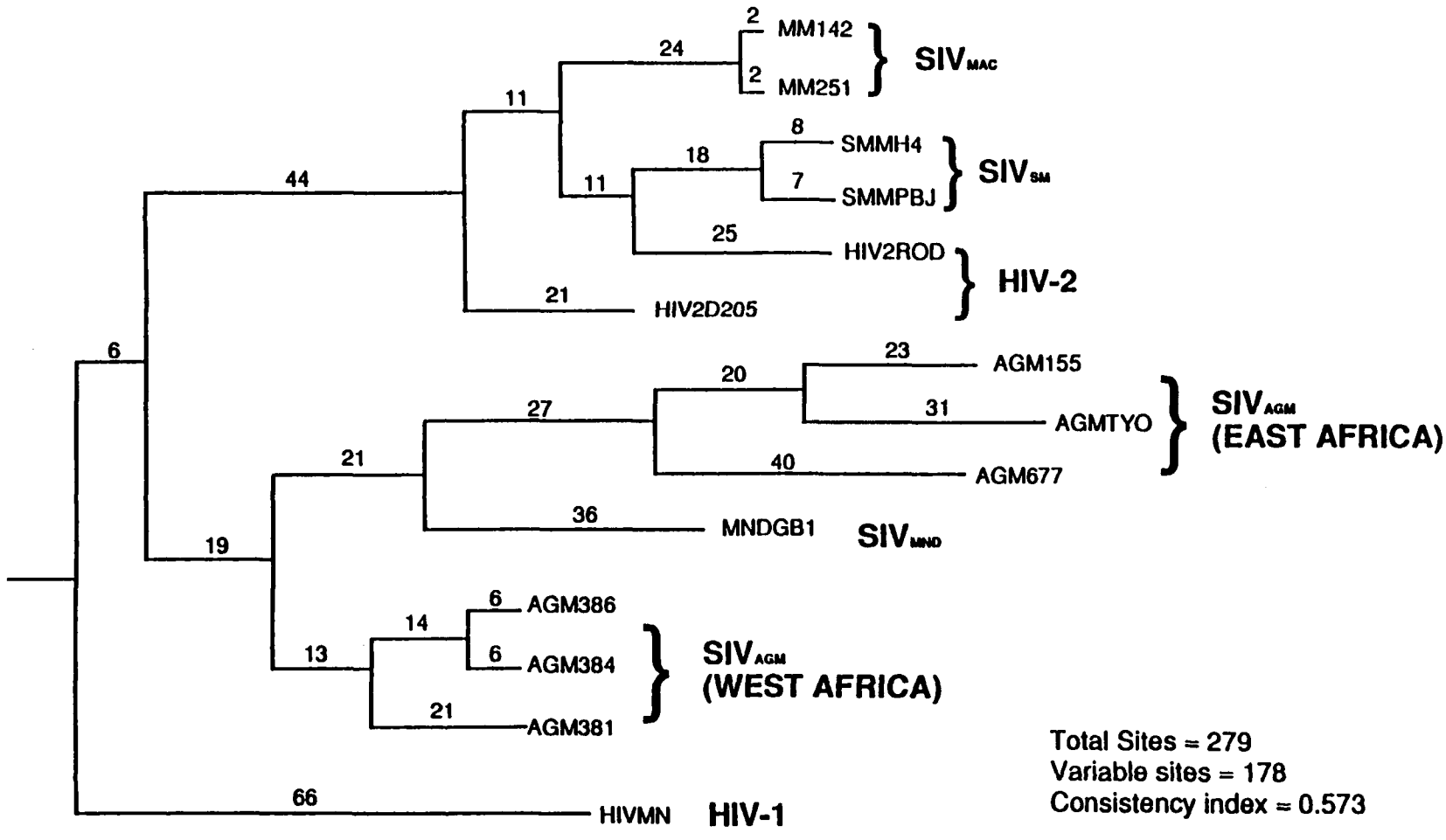


Figure 41. Phylogenetic tree analysis for human and simian immunodeficiency viruses.

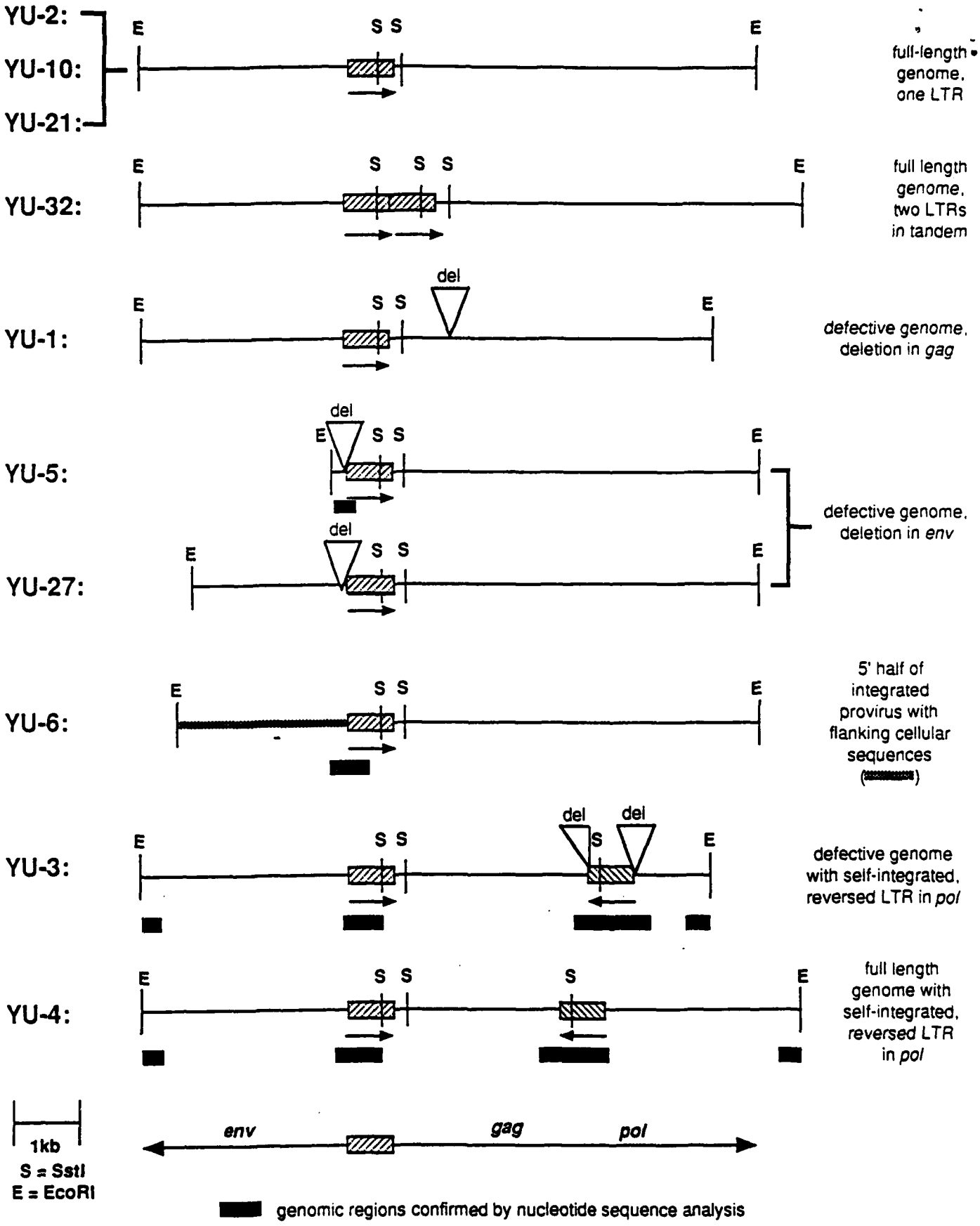


Figure 42. Molecular proviral clones of HIV-1 derived from uncultured brain DNA of a patient with AIDS dementia. Ten HIV-1 proviral clones from a lambda phage genomic library of 8 x 10^6 recombinant phage were identified and mapped. Because the predominant viral forms in this uncultured brain DNA specimen were in unintegrated circular form, a single cutter, EcoRI, was selected as the cloning enzyme. Four of the 10 clones were full-length with either one or two LTRs (YU-2, 10, 21, 32), one clone was partial but integrated into genomic DNA (YU-6), and the remaining 6 clones were defective by virtue of deletions or rearrangements (YU-1, 3, 4, 5, 27).

PREDOMINANT FORM:	YU-3	PCR-B13
	YU-4	PCR-B16
	YU-10	PCR-C6
	YU-21	PCR-C20
	YU-32	
PERCENTAGE:	5/8=62.5%	4/11=36.3%

COMPARISON WITH THE

PREDOMINANT FORM

(bp/bp)

YU-1	1/525=0.19%
YU-2	2/525=0.38%
YU-27	2/525=0.38%
PCR-A7	1/525=0.19%(deletion)
PCR-B7	3/525=0.57%
PCR-B8	3/525=0.57%(2 bp deletion)
PCR-B9	6/525=1.14%(1 stop codon)
PCR-B14	1/525=0.19%
PCR-C14	2/525=0.38%(1 stop codon)
PCR-C19	2/525=0.38%

Figure 43. Summary of envelope nucleotide sequences from eight HIV-1 proviruses in lambda (Yu-1, 2, 3, 4, 10, 21, 27, and 32) and from eleven PCR derived envelope fragments (PCR-A7, B7, B8, B9, B13, B14, B16, C6, C14, C19, C20). All clones were derived from uncultured brain DNA (see text).

	bp/bp	NUCLEOTIDE	AMINO ACID
YU-1:	1/525=0.19%	GAT--AAT	(Asp--Asn)
YU-2:	2/525=0.38%	AAG--AAT	(Lys--Asn)
YU-27:	2/525=0.38%	GTA--GTC	(Val--Val)
		TCT--GCT	(Ser--Ala)
PCR-A7:	1/525=0.19%	TGG--*GG	(Trp-1bp deletion)
PCR-B7:	3/525=0.57%	CTA--CCA	(Leu--Pro)
		AGA--AGG	(Arg--Arg)
		ACT--ACG	(Thr--Thr)
PCR-B8:	3/525=0.57%	CTA--TTA	(Leu--Leu)
		ACA--A**	(Thr--2bp deletion)
PCR-B9:	6/525=1.14%	GAG--GAA	(Glu--Glu)
		ACA--ACG	(Thr--Thr)
		AAT--GAT	(Asn--Asp)
		AAG--AAT	(Lys--Asn)
		ACT--ACC	(Thr--Thr)
		TGG--TGA	(Trp--Stop)
PCR-B14:	1/525=0.19%	ATA--ACA	(Ile--Thr)
PCR-C14:	2/525=0.38%	ACT--ACC	(Thr--Thr)
		TGG--TGA	(Trp--Stop)
PCR-C19:	2/525=0.38%	CTA--TTA	(Leu--Leu)
		TAC--TGC	(Tyr--Cys)

Figure 44. Comparison of nucleotide and amino acid sequences of lambda and PCR derived HIV-1 clones in relation to the predominant (consensus) sequences (see Figure 43). All clones were derived from uncultured brain DNA (see text).

APPENDIX B: PUBLICATIONS BASED ON DAMD17-87-C-7038 CONTRACT WORK

Starcich, B.R., B.H. Hahn, **G.M. SHAW**, P. McNeely, S. Modrow, H. Wolf, W. Parks, E. Parks, S. Josephs, R.C. Gallo and F. Wong-Staal: Identification and characterization of conserved and divergent regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. Cell 45:637-648, 1986.

Hahn, B.H., **G.M. SHAW**, M.E. Taylor, R.R. Redfield, P.D. Markham, S.Z. Salahuddin, F. Wong-Staal, R.C. Gallo, E. Parks, W. Parks: Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science 232:1548-1553, 1986.

Modrow, S., B.H. Hahn, **G.M. SHAW**, R.C. Gallo, F. Wong-Staal, H. Wolf: Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: Prediction of antigenic epitopes in conserved and variable regions. J. Virology 61:570-578, 1987.

Ragni, M.V., Urbach, A.H., Taylor, S., Claassen, D., Gupta, P., Lewis, J.H., Ho, D.D., and **SHAW, G.M.**: Isolation of human immunodeficiency virus and detection of HIV DNA sequences in the brain of an ELISA antibody-negative child with Acquired Immune Deficiency Syndrome and progressive encephalopathy. J. Pediatrics, 110 (6):892-894, 1987.

Hahn, B.H., L.I. Kong, S-W Lee, P. Kumar, M.E. Taylor, S.K. Arya, and **G.M. SHAW**: Relation of HTLV-4 to simian and human immunodeficiency-associated viruses. Nature 330:184-186, 1987.

Hahn, B.H., Kumar, P., Taylor, M.E., Arya, S.K., and **SHAW, G.M.**: Molecular cloning and nucleotide sequence analysis of HTLV-4: Comparison to other human and simian immunodeficiency-associated viruses. In: Human Retroviruses, Cancer and AIDS, UCLA Symposia on Molecular and Cellular Biology, 71:57-74, 1988.

Gibbons, J., Saag, M.S., Parks, W., Parks, E., Hahn, B.H., and **SHAW, G.M.**: Genomic diversity of HIV *in vivo* and *in vitro*. In: Human Retroviruses, Cancer and AIDS, UCLA Symposia on Molecular and Cellular Biology, 71:75-114, 1988.

Kappes, J.C., Morrow, C.D., Lee, S-W., Jameson, B.A., Kent, S.B.H., Hood, L.E., **SHAW, G.M.**, Hahn, B.H.: Identification of a novel retroviral gene unique to HIV-2 and SIV_{MAC}. J. Virology 62:3501-3505, 1988.

Weiss, S.H., Goedert, J.J., Gartner, S., Popovic, M., Waters, D., Markham, P., Veronese, F.D.M., Gail, M.H., Barkley, W.E., Gibbons, J., Gill, F.A., Leuther, M., **SHAW, G.M.**, Gallo, R.C., Blattner, W.A.: Risk of human immunodeficiency virus (HIV-1) infection among laboratory workers. Science 239:68-71, 1988.

SHAW, G.M., Wong-Staal, F., Gallo, R.C.: Etiology of AIDS: Virology, Molecular Biology, and Evolution of Human Immunodeficiency Viruses. In: The Acquired Immune Deficiency Syndrome. Edited by V.T. DeVita, S. Hellman, and S.A. Rosenberg. J.B. Lippencott Co., Philadelphia, pp. 11-31, 1988.

Saag, M.S., Hahn, B.H., Gibbons, J., Li, Y., Parks, E.S., Parks, W.P., **G.M. SHAW**: Extensive variation of HIV-1 *in vivo*. Nature 334:440-444, 1988.

Fisher, A.G., Ensoli, B., Looney D., Rose, A., Gallo, R.C., Saag, M.S., **SHAW, G.M.**, Hahn, B.H., Wong-Staal, F.: Biologically diverse molecular variants within a single HIV-1 isolate. Nature 334:444-447, 1988.

Kong, L.I., Lee, S-W., Kappes, J., Parkin, J., Decker, D., Hoxie, J., Hahn, B.H., **SHAW, G.M.**: West African HIV-2 related human retrovirus with attenuated cytopathicity. Science 240:1525-1529, 1988.

Lee, S.-W., Kappes, J.C., Decker, W.D., Park, J., Morrow, C.D., Su, S, Kong, L.I., Derosiers, R.C., **SHAW, G.M.**, and Hahn, B.H.: Studies on the function of vpx in HIV-2. In: Retroviruses of Human AIDS and Related Animal Diseases, 3e Colloque Des Cent Gardes, pp. 11-17, 1988.

Kumar, P., Hui, H., Kappes, J.C., Haggarty, B.S., Hoxie, J.A., Arya, A.K., **SHAW, G.M.** and Hahn, B.H.: Molecular characterization of an attenuated isolate of HIV-2. J. Virology, 64:890-901, 1990.

Allan, J.S., Short, M., Taylor, M.E., Su, S., Hirsch, V.M., Johnson, P.R., **SHAW, G.M.**, and Hahn, B.H. Natural infection of West African Green Monkeys with a highly divergent substrain of SIV_{AGM}. In preparation.

Li, Y., Price, R.W., Hahn, B.H and **SHAW, G.M.** Molecular cloning and analysis of replication competent HIV-1 proviruses from uncultured human brain in AIDS dementia complex. In preparation.

APPENDIX C: PERSONNEL EMPLOYED ON CONTRACT DAMD 17-87-C-7038

Name/Degree	Position on Contract	Present Position
George M. Shaw, M.D., Ph.D.	Principal Investigator	Principal Investigator
Beatrice H. Hahn, M.D.	Co-Investigator	Co-Investigator
Sajal Ghosh, Ph.D.	Postdoctoral Fellow	Postdoctoral Fellow
Theda Jackson	Laboratory Technician	Laboratory Technician
Prasanna Kumar, Ph.D.	Postdoctoral Fellow	Research Associate
Yuexia Li, B.S.	Predoctoral Research Fellow	Predoctoral Research Fellow
Maria Taylor, M.S.	Research Assistant	Research Assistant
Diccie Dickie	Laboratory Technician	Transfer
Carol Hartline, B.S.	Research Assistant	Transfer
Lilly Ing Kong, D.V.M.	Research Associate	Transfer
Shei-Wen Lee, D.V.M.	Postdoctoral Fellow	Transfer
Tashana Thomas	Laboratory Technician	Transfer

DISTRIBUTION LIST

4 copies	Director Walter Reed Army Institute of Research ATTN: SGRD-UWZ-C Washington, DC 20307-5100
1 copy	Commander US Army Medical Research and Development Command ATTN: SGRD-RMI-S Fort Detrick, Frederick, Maryland 21702-5012
2 copies	Defense Technical Information Center (DTIC) ATTN: DTIC-DDAC Cameron Station Alexandria, VA 22304-6145
1 copy	Dean School of Medicine Uniformed Services University of the Health Sciences 4301 Jones Bridge Road Bethesda, MD 20814-4799
1 copy	Commandant Academy of Health Sciences, US Army ATTN: AHS-CDM Fort Sam Houston, TX 78234-6100